

Twentieth Army Science Conference

Twentieth Army Science Conference

Award Winning Papers



Editors

Richard Chait, Catherine Kominos,
Michael S. Shur, Michael Stroscio
& James J. Valdes

World Scientific

Chait, Kominos,
Shur, Stroscio & Valdes



Twentieth Army Science Conference

Award Winning Papers

20010515 034

AQ401- 08-1399

Twentieth Army Science Conference

Award Winning Papers

Norfolk, Virginia, USA 24 – 27 June 1996

Editors

Richard Chait

Army Research Office of the Assistant Secretary of the Army

Catherine Kominos

Army Research Office of the Assistant Secretary of the Army

Michael S. Shur

Rensselaer Polytechnic Institute

Michael Strosio

US Army Research Office

James J. Valdes

US Army Edgewood Research, Development and Engineering Center



World Scientific

Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

TWENTIETH ARMY SCIENCE CONFERENCE: AWARD WINNING PAPERS

Copyright © 1997 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-02-3178-4

Printed in Singapore by Uto-Print

PREFACE

This volume contains the twenty-one prize-winning papers presented at the Twentieth United States Army Science Conference which was held in Norfolk, Virginia during 24–27 June 1996 under the sponsorship of the Assistant Secretary of the Army for Research, Development and Acquisition. The theme for the Twentieth Army Science Conference was Science and Technology for Force XXI — the Force of the Twenty-First Century. At the conference, original scientific and technical papers — written by U. S. Army civilian and military scientists and engineers — addressed the role of science and technology in providing a competitive edge for Force XXI. Over one hundred and fifty original papers were presented at the Twentieth Army Science Conference in the following ten key technical areas:

- Advanced Materials and Manufacturing including Structures and Energetics;
- Microelectronics and Photonics;
- Sensors and Information Processing including Communications;
- High-Performance Computing and Simulation including Modeling, Displays, Artificial Intelligence, and Virtual Reality;
- Advanced Propulsion Technologies including Mobility and Lethality;
- Power Generation, Storage and Conditioning including Directed Energy;
- Biological, Chemical and Nuclear Defense;
- Life, Medical and Behavioral Sciences including Individual Sustainability and Soldier-System Interface;
- Environmental Sciences and Geosciences including Atmospheric, Space and Environmental Protection;
- Engineering Sciences including Robotics, Mechanics, Fluid Dynamics and Survivability.

Preceding Page's Blank

Each of the twenty-one award-winning papers in this volume has been reviewed by expert peer reviewers from the community at large. It is indeed a great pleasure to acknowledge all of these reviewers for their advice. Many of these distinguished scientists and engineers have granted permission for the editors to acknowledge their services publicly; accordingly, the editors express their sincere appreciation to the following scientists and engineers: Professor Romesh C. Batra, Virginia Polytechnic Institute and State University; Professor Ted Belytschko, Northwestern University; Professor Terrill Cool, Cornell University; Professor Yuki Horie, North Carolina State University; Dr. Angray S. Kang, Scripps Research Institute; Professor Narayanan Komerath, Georgia Institute of Technology; Professor Alan R. Mickelson, University of Colorado; Professor Fred H. Pollak, Brooklyn College of the City University of New York; Professor Donald Thompson, Oklahoma State University; Professor John Walsh, Dartmouth College; Professor Ronald Williams, University of Virginia at Charlottesville, and Dr. Steve Lang of Lockheed-Martin Company.

Through this volume, the editors hope to highlight the importance of Army Science and Technology.

Richard Chait, U.S. Army
Catherine Kominos, U.S. Army
Michael Shur, Rensselaer Polytechnic Institute
Michael A. Stroschio, U.S. Army
James J. Valdes, U.S. Army

CONTENTS

Preface	v
---------------	---

Physics

A Lightweight Electron Beam Focusing Structure for Missile Radars	3
<i>Herbert A. Leupold</i>	

Machine Aided Search: Results of Human Performance Testing Using Automatic	8
Target Recognition and Second Generation Forward Looking Infrared Sensors	
<i>D. A. Reago, Jr. and W. C. Gercken</i>	

High-Resolution Acoustic Direction-Finding Algorithm to Detect and Track	15
Ground Vehicles	
<i>Tien Pham, Brian Sadler, Manfai Fong and Donald Messer</i>	

Nonlinear and Quantum Optics of Multicomponent Media.....	21
<i>Michael E. Crenshaw and Charles M. Bowden</i>	

Mathematical and Computer Sciences

Parallel Finite Element Computation of Missile Flow Fields	29
<i>W. Sturek, S. Ray, S. Aliabadi, C. Waters and T. Tezduyar</i>	
Building Simulations for Virtual Environments and Prototyping.....	34
<i>Ming C. Lin</i>	

Chemical and Biological Sciences

Theoretical Determination of Potential Hazards in the Handling of CK and	41
AC Munitions	
<i>Betsy M. Rice, Sharmila V. Pai and Cary F. Chabalowski</i>	

Carboxylesterase: Regulatory Control and Peptide-Induced Secretion of an Endogenous Scavenger for Organophosphorus Agents	46
<i>Donald M. Maxwell, Kenneth D. Lanclos and Hendrik P. Benschop</i>	
Measuring Visual Resolution in the Contrast Domain: The Small Letter Contrast Test.....	51
<i>Jeff Rabin</i>	
Production of Recombinant Antibodies for Biosensor Applications	56
<i>Peter A. Emanuel, James J. Valdes and Mohyee E. Eldefrawi</i>	
A Protein Engineering Approach to Designing Staphylococcal Enterotoxin Vaccines	61
<i>Robert G. Ulrich</i>	
Molecular Beam Mass Spectrometric and Modeling Studies of Neat and NH_3	67
Doped Low Pressure $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ Flames: Formation and Consumption of NO <i>R. C. Sausa, W. R. Anderson, G. Singh and G. W. Lemire</i>	
Engineering and Environmental Sciences	
Laser Velocimetry and Doppler Global Velocimetry Measurements of Velocity	75
Near the Empennage of a Small-Scale Helicopter Model <i>Susan Althoff Gorton, James F. Meyers and John D. Berry</i>	
Modeling of the Mixing/Combustion Process in a Quiescent Chamber D. I.	80
Diesel Engine <i>P. Schihl, W. Bryzik, E. Schwarz and E. Danielson</i>	
The Fabrication and High-Strain-Rate Properties of Hot-Explosively-Compacted	93
Tungsten-Titanium Alloys <i>Laszlo J. Kecskes and Ian W. Hall</i>	
Concrete Constitutive Modeling in High Velocity Penetration Analysis	98
<i>V. M. Gold, G. C. Vradis and J. C. Pearson</i>	
Source Characterization Modeling for Demil Operations.....	103
<i>C. A. Biloft, E. S. Oran, J. P. Boris, C. A. Lind and W. J. Mitchell</i>	

Electronics

Strain Engineered Semiconductor Heterostructures for Novel Optoelectronic Devices..... 111

Paul H. Shen, Jagadeesh Pamulapati, Michael Wraback, Weimin Zhou,

Monica Taysing-Lara and Mitra Dutta

New Light Coupling Scheme and Quantum Transition Noise of Quantum 116
Well Infrared Photodetectors

K. K. Choi, W. H. Chang, C. J. Chen and D. C. Tsui

Advances in Quasi-Optical Power Combiners Provide Path to Radar and 121
Communications Enhancements

J. F. Harvey, M. B. Steer, J.W. Mink, H.-S. Hwang, T. W. Nuteson

and A. C. Paolella

Multi-Process Synthesis of Novel Ferroelectric Oxide Ceramic Composites for 126
Use in Phased Array Antennas

L. C. Sengupta, E. Ngo, S. Sengupta and S. Stowell

List of Contributors..... 135

Twentieth Army Science Conference

Award Winning Papers

Physics

A Lightweight Electron Beam Focusing Structure for Missile Radars

Herbert A. Leupold
U. S. Army Research Laboratory
Fort Monmouth, NJ 07703-5601

Abstract

A 15 kg permanent magnet electron-beam focusing structure has been invented, designed and built to replace a 55 kg electromagnet in Generic Missile Application Program (GMP) missiles, thereby affording considerable increase in range, payload and mobility. Unlike the electromagnet the structure does not entail the inconvenience of electric power sources, massive copper coils or heavy paraphernalia.

The permanent magnet is in the form of a solenoid from which flux leakage is largely eliminated and which furnishes a uniform axial electron-guide field of 2.4 kOe over a 23 cm distance in a tube 3.5 cm in diameter. A flux free space at one end of the structure accommodates a field sensitive electron-beam source. All specifications were met by magnetic compensation of field distortions caused by access holes to the structural interior for leads and wave guides. A prototype was built and tested on a beam stick which revealed a 98% beam transmission on the first try, indicating easy fulfillment of the required 95%.

The transfer of funds from an industrial concern (Martin-Marietta) to a military laboratory (ARL) under the terms of an CRDA effort specifying the project is believed to be unprecedented.

1. Introduction

Most high powered radiation sources employed by the military derive their energies from electron beams that require strong magnetic fields to guide and focus them. In many cases the required fields are solenoidal and are traditionally furnished by electrical coils. These are cumbersome, require power supplies and cooling coils, and are far too massive and bulky for many airborne and ballistic devices. Until relatively recently, permanent magnets were not a viable alternative to electric coils because traditional magnetic materials such as ALNICO's do not possess sufficient coercivity to maintain their magnetization in the face of high demagnetizing fields encountered in the required structures.

The magnetic group at ARL recognized that the advent of high-energy rare earth magnets presented a promising opportunity for the design of a plethora of theretofore impractical devices that afforded high power, compactness, light weight and sometimes custom features such as minimum stray flux and shielded chambers to house field-sensitive components. Accordingly, the magnetics group at ARL developed the novel design techniques needed to efficiently exploit the high energy materials to make of them a revolutionary difference in kind rather than an incremental improvement of degree. This work resulted in an extensive design and patent inventory so that when ARL was solicited by the Martin-Marietta Company to design a magnet for a radar tube, ARL had necessary technology in place. The tube was for a GMP missile that would be far too heavy if it used a conventional field source.

2. Specifications and Design

The magnet was to furnish an axial field of 2.4 kOe in a cylindrical space 27 cm long and 3.5 cm in diameter. The field was to be uniform to within 7%. At one end of this space was to be a smaller, field free

cylindrical chamber to accommodate a field-sensitive electron gun. Access needed at the gun end of the structure would occupy almost the entire cross section of the working space, and a rectangular aperture near the beam collector at the other end was required for admittance of a wave guide. Mass, bulk and flux leakage were to be minimal. The field is supplied by a cylindrical shell magnet magnetized in the direction of the axis and is conducted into the central working space by iron discs, capping the ends of the shell as in figure 1. We assume that all of the supplied flux is confined to the working space and then deduce the conditions that make it possible. We note that by conservation of flux, Φ ,

$$\Phi_m = \Phi_w = B_m A_m = B_w A_w \quad (1)$$

where B_m and B_w are the flux densities in the magnet and working space respectively and A_m and A_w are the respective cross sectional areas of magnet and working space. But in Gaussian units $B_w = H_w$ so that

$$A_m = H_w A_w / B_m \quad (2)$$

For a perfectly rigid magnet such as rare earth permanent magnets the internal flux density is given by

$$B_m = -H_m + B_r = -H_w + B_r \quad (3)$$

where B_r is the magnetic remanence. Application of Ampere's Law to the cylindrical wall of the chamber tells us that $H_m = H_w$ and we have for the cross sectional area of the magnet

$$A_m = H_w A_w / [B_r - |H_w|] = A_w / [B_r / H_w - 1] \quad (4)$$

where all the quantities on the right are specified and A_m is determined.

The dimensions and form of the magnetic arrangement that must clad the supply magnet and its poles to confine the flux can be deduced from consideration of the circuital form of Ampere's law:

$$\oint \vec{H} \cdot d\vec{l} = 4\pi I \quad (5)$$

where I is any electric current linked by the loop of the line integral. There are no electric currents I in our configuration so the line integral must be zero. We choose for the path of integration ABCD of figure 2 where

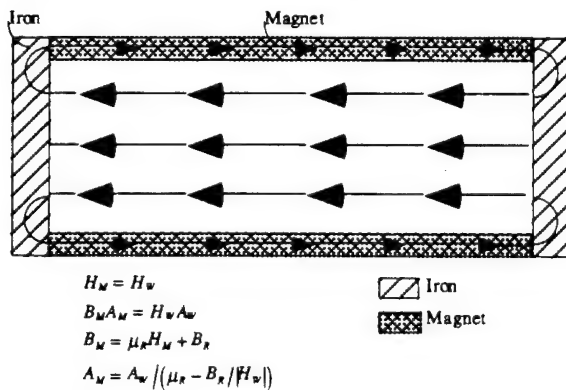


Figure 1. Supply magnet, pole pieces and working space. For ideal magnet $\mu_r = 1$. Heavy arrows are flux in cavity. Light arrows are magnetization.

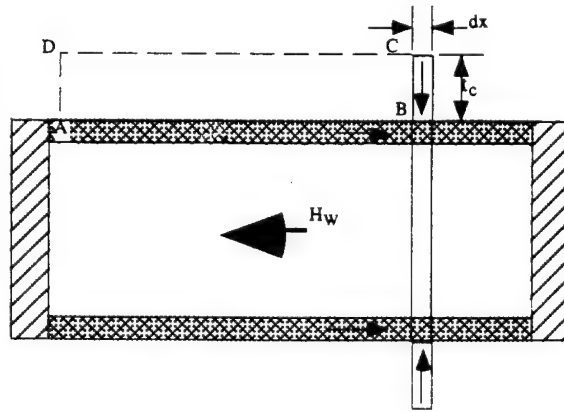


Figure 2. Cladding Magnet Construction

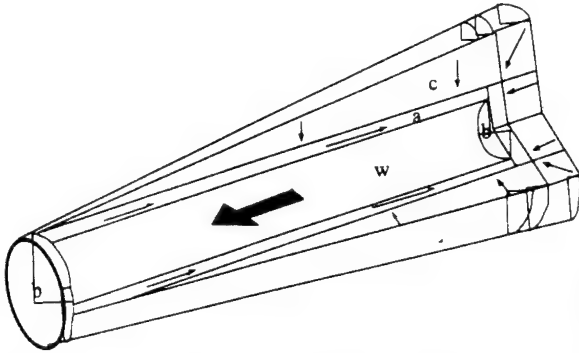


Figure 3. Permanent Magnet Solenoid. (a) supply magnet (b) pole pieces (c) cladding (w) work space

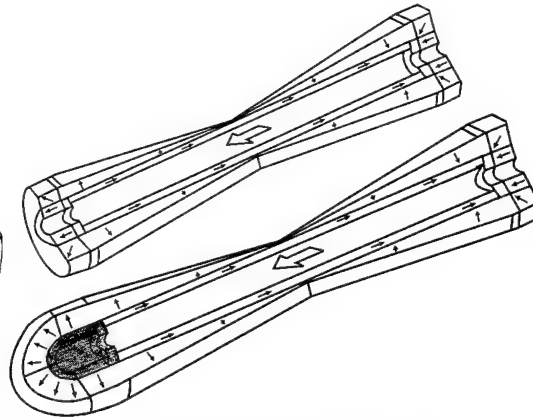


Figure 4. Two permanent magnet solenoids. (A) Without field-free Chamber, (B) with field-free electron gun chamber on left end.

C is any point on the outer surface of the cladding. Since, by specification, the field over the path CDA must be zero, it follows that integration over the remainder of the path must also be zero and therefore:

$$\int_A^B \vec{H} \cdot d\vec{l} = - \int_B^C \vec{H} \cdot d\vec{l} \quad (6)$$

The left side of the equation is just $H_m(AB) = H_w(AB)$ and since by hypothesis no flux is to flow to the exterior via BC, the flux density must be zero there. From (3) we see that H must be equal to B_r when B is zero and if t_c is the thickness (BC) of the cladding at BC, (6) becomes

$$H_w(AB) = B_r t_c \rightarrow t_c = H_w(AB) / B_r \quad (7)$$

and t_c varies linearly with distance AB and a hollow conical cladding with radial magnetization results. The cladding at the end of the structure must be a disc of the same thickness $t_c(\max)$ of the hollow conical side cladding there as shown in figure 3.

The magnet of figure 3 is not entirely satisfactory because of the still excessive mass that specified fields require. At ARL it was noted that mass can be reduced to less than half if we choose our zero potential to be at mid-length of the chamber. Then the magnetization of the two resulting cones must point outward in the cone to the left of the zero potential plane and inward in the right cone. The total cladding mass would then be only about 40% of that for a structure with zero potential at one of the ends. Such a structure is shown in figure 4(A).

The housing of the electron gun at one of the ends must be field-free within a small specified tolerance and yet must be in close proximity to the working space where the field is full value. Since only a small aperture is needed for the electron beam to traverse from gun to working space, a hole penetrating the pole piece at the end of the working chamber will serve this purpose while the pole piece provides a magnetic shield for the gun. The gun chamber as a whole must be encased by iron and by permanent magnetic cladding of thickness equal to that of the maximum thickness of the conical cladding as in figure 4(B). If this is done the optimal placement of the zero potential plane is no longer at the central plane of the working space. The optimal position is easily determined if the total required cladding mass is stated in terms of the location of the zero plane and the resulting expression differentiated with respect to the coordinate X , of that location, and set equal to zero.

Because the required access holes are so large, it is necessary to deal with the field distortions they entail. To accomplish this, the unmodified structure is modeled on a three dimensional computer program and the resulting field mapped everywhere. Deviations of the field from that desired are noted, and appropriate modifications are made in the structure to eliminate them. Experience has furnished us with a variety of methods

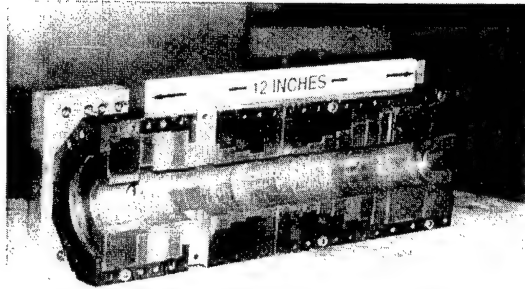


Figure 5. Sectional details of magnet structure. (A) Partial cladding ring in assembly jig, (B) section of assembled magnet.

to accomplish this. Some are relatively sophisticated such as the appropriate placement of small dipole magnets to cancel the major Fourier components of the deviations. Others are cruder, partially computational and partially trial and error such as variation in dimensions, direction and strength of magnetization.

To eliminate small dips or rises in the axial field profile it is sometimes useful to place an axially magnetized ring of appropriate dimensions around the area of the irregularity so that the field of the ring cancels it. Radially magnetized rings are also useful for the cancellation of more complex irregularities and also for making desired field changes more abrupt. An example is the transit from the gun chamber to the working space where the field must change to 2.5 kOe in as short a distance as possible to maintain close to zero field at the gun and to maintain close focus of the electron beam after it leaves the gun. With these techniques we incorporate the appropriate alterations to cancel the perturbations revealed by the first computer field plot and then replot. Additional alterations are made in the design and by successive approximations the desired field profile is attained.

3. Construction and Evaluation

The prototypical structure was built by Magnet Sales Corporation and is pictured in figure 5. The clam shell arrangement affords easy interior access for tube insertion and measurement. The cladding is a stepped arrangement rather than a continuous cone for convenience of manufacture. Computer analysis shows that this does not significantly affect the field. The magnets are of Nb-Fe-B compound with a remanence of 13 kG and are made in the form of annular rings assembled from wedge shaped segments. The rings are held together and to each other with epoxy and a metal jig for safety.

Field measurements on and off axis show that the structure does indeed meet all specifications as shown in figure 6. The measured field is actually higher than predicted because magnetic material of higher remanence was used in construction than assumed in computation. When this difference is taken into account there is a near-perfect agreement. The field in the gun chamber nowhere exceeds 25 gauss, well below the specified maximum of 50 gauss. The axial field showed a deviation from uniformity no greater than 3% compared to 7% allowable in the interactive region.

Finally, an electron beam source and beam stick was inserted to test beam focus. Transmission from source to focus collector was found to be 98% on the first attempt, well above the 95% desired, so no further structural modification or tuning was necessary. This is accomplished with a mass of 15 kg compared to the 55 kg of the original electromagnet. See figure 7 for comparison.

4. Conclusion

The invention, design, manufacture and testing of a permanent magnet solenoid were successfully completed at the request of Martin-Marietta to make a high energy traveling wave tube suitable for CORPSAM use. This was done under a CRDA agreement under which funds were transferred from Martin-Marietta to ARL. The direction of transfer is believed to be unprecedented. This work augurs similar future projects in which the performance of much of the present day high power radiation equipment will be greatly enhanced.

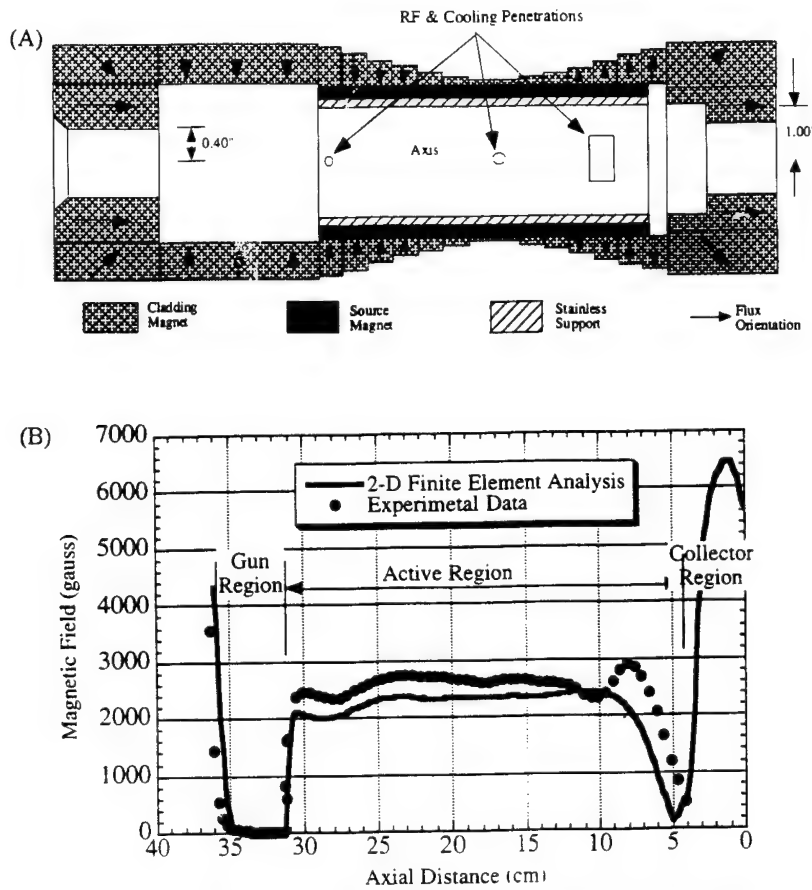


Figure 6. Schematic of (A) Final Structure and (B) Field Profile

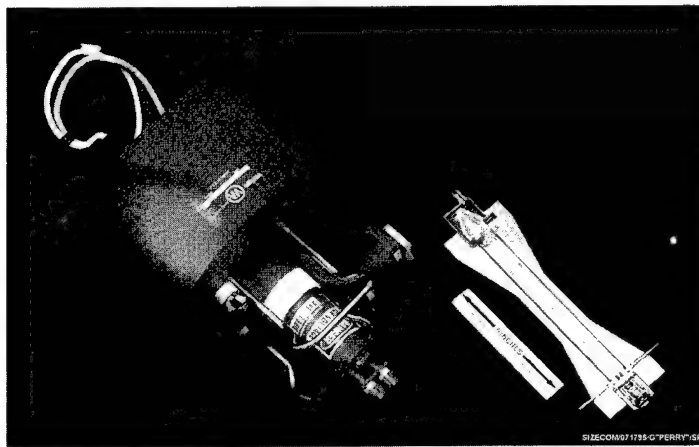


Figure 7. Size Comparison Between Electromagnetic (left) and Permanent Magnet Focusing Structures

Acknowledgments

The author thanks Dr. E. Potenziani II for his help in the computation and analysis of all the designs and Ms. O. Blum and Dr. A. Tilak for their help in preparing the manuscript.

Machine Aided Search: Results of Human Performance Testing Using Automatic Target Recognition and Second Generation Forward Looking Infrared Sensors

D. A. Reago, Jr.* and W.C. Gercken
U.S. Army CECOM Night Vision and Electronic Sensors Directorate
Ft. Belvoir, Va. 22060

Abstract

Part task mission simulations were conducted by the CECOM Night Vision and Electronic Sensors Directorate on the effectiveness of trained observers performing target search with a 2nd Generation Forward Looking Infrared (FLIR) sensor enhanced with Automatic Target Recognition (ATR) cues. The purpose of this experiment was to evaluate the performance impacts of introducing ATR information into a reconnaissance task applicable to attack helicopters or real-time unmanned aerial vehicles (UAV's). A manual system mode that required observers to visually search for targets within a field of regard was compared to an aided system in which the observers were presented a series of cues derived from ATR processing of the FLIR imagery. The aided man-machine interface was optimized for use with the ATR information and was similar to the one used in the Multi-Sensor Aided Targeting-Air program. The results showed a significant improvement in detection rates in the aided search mode versus the manual search mode. They also demonstrate that the *overall* number of false alarms reported is lower during aided search, even accounting for those false alarms generated by the ATR and presented to the observer. The results confirm that if ATR's achieve performance levels in actual fielded systems comparable to those currently being measured in technology demonstration programs, then significant improvements in overall warfighting capability can be expected. They also demonstrate the need for an optimized system concept and machine interface to fully exploit the advantages of ATR information during time compressed searches.

1. Introduction

Automatic or Aided Target Recognition (ATR) has been recognized as a highly desirable capability that will significantly enhance the lethality and survivability of U.S. Armed forces. Several major weapon systems are being developed or deployed that depend, in part, on ATR capabilities to meet their operational requirements. These include the AH-64D Longbow Apache, the RAH-66 Comanche, and the Brilliant Anti-Tank (BAT) submunition. Future systems to employ ATR may include the family of armored vehicles (M-1/M-2) as well as synthetic aperture radar (SAR) and FLIR sensors flying on UAV's. With the exception of the BAT seeker, the rest of these weapon systems employ the ATR to enhance the speed and accuracy of the overall search process, which includes a human operator reviewing the ATR results and making final targeting decisions. Figure 1 illustrates the cumulative sensor, processor and human operator *aided search* process.

The Department of Defense has and continues to make a significant investment in the development of practical algorithms and signal processors for ATR. Recent Army Science and Technology program examples include the Multi-Sensor Aided Targeting-Air (MSAT-Air) ATD (Lloyd et al., 1995) and the SAR Target Recognition and Location (STARLOS) system (Rodkey et al., 1995). These and other programs have demonstrated that ATR performance meeting weapon system requirements is achievable in low to moderate clutter using advanced high resolution sensors and ATR algorithms based on template matching and neural network approaches (Bohan et al., 1996).

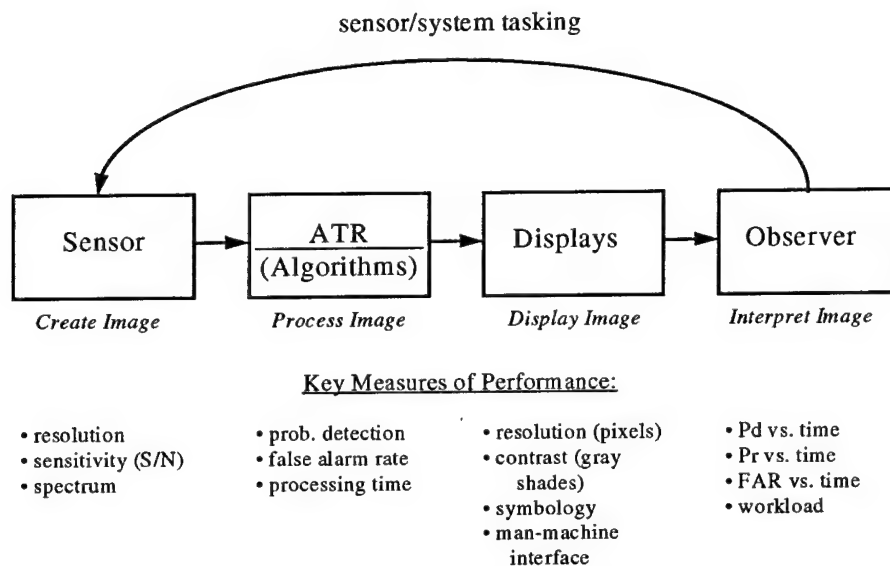


Figure 1. Man-in-the-loop aided search process

Despite these successes, there remains a relatively limited amount of data concerning the cumulative man-machine (aided) performance during search. Notable exceptions include cued search performance experiments performed at the Night Vision Laboratory using terrain board (synthetic FLIR) imagery (Blecha et al., 1991) and recent experiments conducted as part of the Hunter Sensor Suite ATD (O’Kane et al., 1996) using real 2nd Generation FLIR imagery and simulated ATR results. Blecha et al. demonstrated that at some threshold in ATR performance (Probability of Detection/False Alarm Rate) machine aided search exceeded purely manual search. O’Kane et. al demonstrated that with realistic imagery and comparatively good ATR performance, significant improvements in overall search performance were possible. Key features of both these experiments were a relatively long search time and a man-machine interface that simulates the normal display of a single FLIR field of view within an overall search field of regard (with added cues), in a manner similar to traditional manual search¹.

The current effort extends our understanding of aided search performance to highly time constrained searches that might be found in an reconnaissance/attack helicopter or in a real-time UAV image processing ground station. Aided search, using a man-machine interface designed to speed the interpretation of ATR results, was measured and compared to the performance of purely manual search. The 2nd generation FLIR imagery for the experiments used was representative of modern tactical systems. ATR results were obtained by processing the recorded imagery through the MSAT-Air ATR algorithm suite. The results confirm previous conclusions (Nicholl, 1994) that the detection performance advantages from aided search are greatest for short times, and that human observers tend to find false detections (false alarms) at a constant rate. The results also demonstrate the importance of a suitable man-machine interface to enhance the speed benefits of ATR-aided search.

¹In traditional manual search, the observer directs the sensor line of sight through a field of regard. The ATR results in the referenced experiments were directly superimposed over the FLIR imagery within the search field of regard. Observers were required to view the entire field of regard in order to see the ATR results.

2. Description of Experiment

For this experiment, a part task simulator (SUN workstation) was used to represent the sensor, ATR processor and display system. The sensor was a high performance 2nd Generation FLIR representative of the state-of-the-art for airborne systems. Sensor imagery was digitally recorded from field collections and used to form the search field of regard for the experimental trials. The sensor imagery was processed through the MSAT-Air ATR algorithm suite to generate target detections and false alarm locations. Two aided configurations with different Pd and FAR were evaluated. The SUN workstation display format was programmed to simulate the man-machine interface found in the MSAT-Air aircraft. Display sizes and observer to display distance were adjusted to correlate with the airborne implementation. Observers interacted with the display through the use of a mouse, pointing and clicking on simulated display buttons and controls. An actual fielded system would use mechanical controls and touch screens, and would be expected to produce faster results.

The general experimental scenario was to perform a target search from the perspective of a reconnaissance helicopter pilot. The task presented to the subjects was to reconnoiter a target area for which they have no prior intelligence and to formulate a list of targets, nominally for hand-off to other aircraft assets. Due to survivability and data timeliness constraints, the overall search task was limited to a relatively short duration (less than one minute). Subjects were told in advance of the time constraints, and were kept aware of the elapsed time during the search process. The search imagery was collected in a series of *gimbal scans*². The ATR was assumed to process the imagery in real time as it was collected. (ATR results were precomputed prior to the search experiment. Replay of the imagery during the trials with ATR cues made it appear as if the ATR had performed its target detection process while the scan was being collected).

Figure 2 shows the overall man-machine interface used during the aided mode. The left image display is referred to as the *chip display*. Each ATR result within the field of regard was reduced to an image "chip" and organized with respect to its position in the field of regard (left to right, top to bottom). The right image display is the *frame display*. The frame display shows a single FLIR field of view (FOV) within the overall scan field of regard (FOR). The bar at the bottom of the frame shows the position of the current FOV within the overall FOR. For the example in figure 2, the frame display shows the fourth FOV in an extended scan FOR of five FOV's. The operator can use the buttons found at the bottom of the chip display to peruse the entire chip list (if it should exceed 6 chips) and the buttons at the bottom of the frame display to shift the display FOV within the overall FOR. A target may be confirmed in the chip display by a mouse click within the chip itself; for the frame display a target may be marked by clicking the mouse over the top of the target. Subjects were directed to perform the "confirming" tasks as rapidly as possible (i.e. not to wait until the end of the search trial to make their decisions), to better simulate the tactical mission. Subjects were also asked to review all of the chips prior to reviewing any of the frames. The manual search process was very similar, except that there were no chips displayed or target cues placed in the frame display.

The image set was taken from MSAT-Air data collections performed at Yuma Proving Grounds, Az. As can be seen from figure 2, this desert imagery presents a significant challenge to the observer to detect and recognize targets within a background scene populated by many target-like "bushy" objects. Although the example targets shown here are quite large due to their closeness to the sensor, many of the targets were substantially smaller and harder to detect. The target set consisted of a mix of US and Russian tactical targets and is listed in Table 1. Each FOR was approximately 15° in azimuthal extent and there were an average of three targets per FOR.

² A gimbal scan is a search process which enables the airborne FLIR (in a turret) to be rapidly slewed through a sizable field of regard (45° in azimuth), with the resulting imagery stored in digital memory for recall by the crew. The result is a "strip" of imagery which may be reviewed by the crew in a masked position.

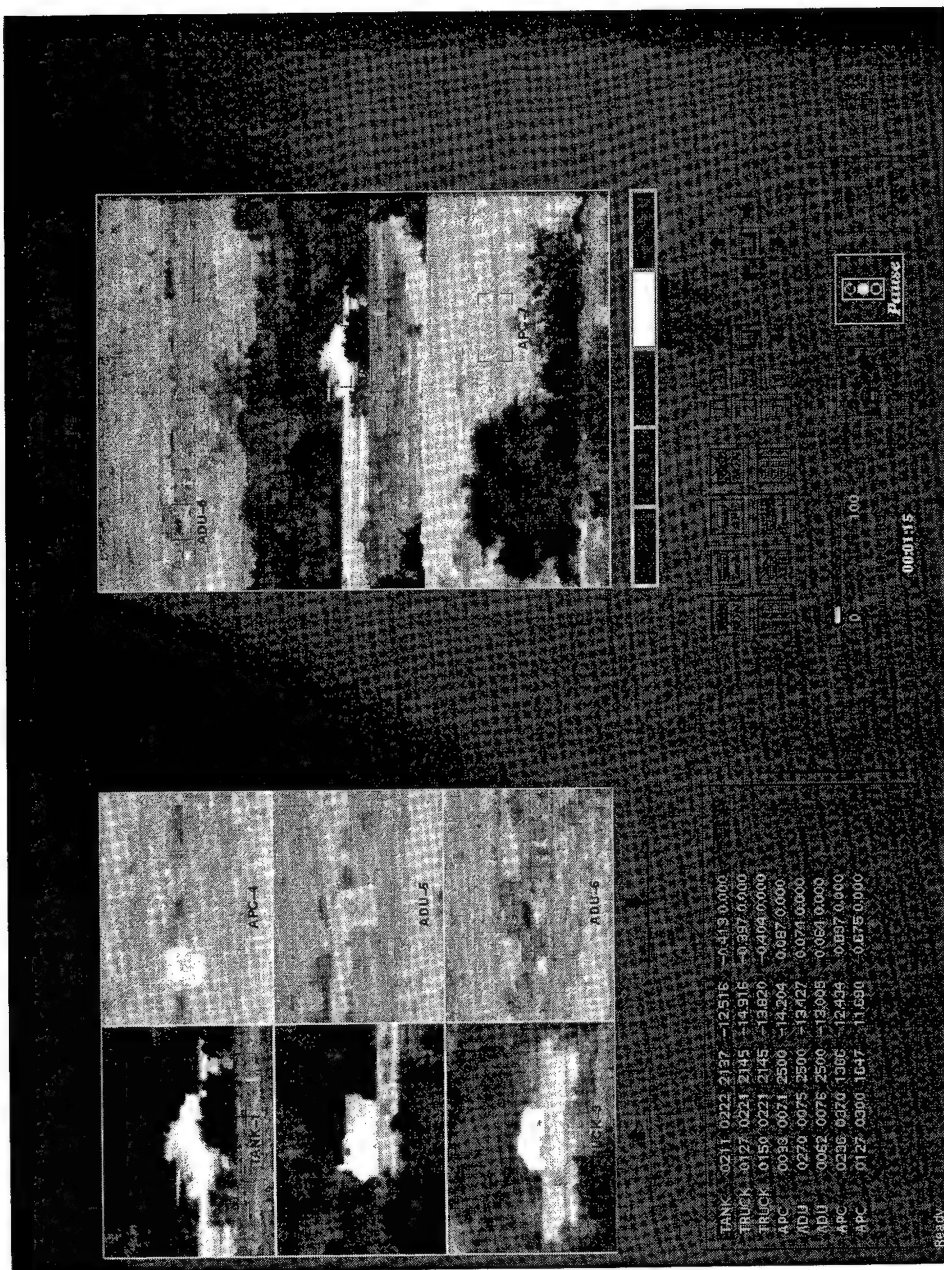


Figure 2. Man-Machine Interface for Search Experiment

Table 1. Search Test Target Set^a

US Targets	Russian Targets
M-60 (Tank)	T-72 (Tank)
M-1 (Tank)	BMP-2 (APC)
M-2 (APC)	ZSU-23 (Air Defense)
M-113 (APC)	2S1 (Air Defense)
HMMWV (Truck)	
M-35 (Truck)	

^aThe target class for each target type is shown in parenthesis.

Two ATR performance settings were used during the trials. The first represents a very high performance system with a Pd of 0.9 and an ATR false alarm rate of 0.06 false alarms/deg². The second represents an ATR operating in a degraded (high false alarm) mode with a Pd of 0.9 and a false alarm rate of 0.35 false alarms/deg². There was little difference in the resulting aided performance in either case, so the results were averaged to improve the statistical power of the experiment. Each target chip was tagged by the ATR with a class labeled (as shown in Table 1), with probability of being correct of 0.7. Incorrect classifications were randomly assigned. To be consistent with standard search doctrine, observers were asked to change the target classification of the chip if they found it to be in error. Trials performed with the ATR are collectively referred to as "aided mode", while those performed without the ATR are referred to as "manual mode."

The same test was performed on two sets of subjects. The first set was a group of 15 pilots with AH-64, RAH-66 and OH-58 experience from the U.S. Army Aviation Center at Ft. Rucker, Alabama ("Ft. Rucker" set). The second was a group of 30 scientists and engineers with thermal imaging experience from the Night Vision and Electronic Sensors Directorate at Ft. Belvoir, Virginia ("NVL" set). Each subject was given a training period which consisted of self-paced familiarization with the controls and displays in both the aided and manual modes, followed by two sample timed searches (one per mode). Each subject was then given a sequence of 6 FOR's to search (3 aided, 3 manual). Each FOR contained 12 FLIR FOV's.³ The ordering of the configurations was balanced across the entire subject set to ensure that there was no bias in the order of presentation of scenes and search modes. All subject inputs were automatically time tagged and logged into a file to permit accurate post test analysis.

3. Test Results And Discussion

Results were compiled for the two observer sets ("NVL" and "Ft. Rucker"). Data from all subjects, scene sequences and ATR configurations were combined to produce a manual and an aided ensemble average for each observer set.⁴ Figure 3 shows the composite probability of detection vs. time for the aided and manual trials. Figure 3 clearly shows the rapid rise in detection probability accompanying the review of the chips in the aided mode, while the manual mode shows a much slower increase in detection probability. By the end of one minute, observers performing manual search were able to detect less than 50% of the targets within the FOR, while nearly 100% of the targets were detected in the aided mode. Of particular interest is the time to detect the first target. Figure 4 illustrates the average time to first target for each observer set. Although there is some difference between the NVL and Ft. Rucker observers in the manual mode detection time, the overall decrease from manual to aided mode is at least a factor of three.

³ For this experiment, the FLIR FOV was 1.57° (vertical) by 1.17° (horizontal). The total FOR was 1.57° (v) by 14° (h).

⁴ Search model formulations normally rely on the concept of an ensemble average over a set of observers and scenes to produce predictable and verifiable results. For example, see Lawson et al. (1978).

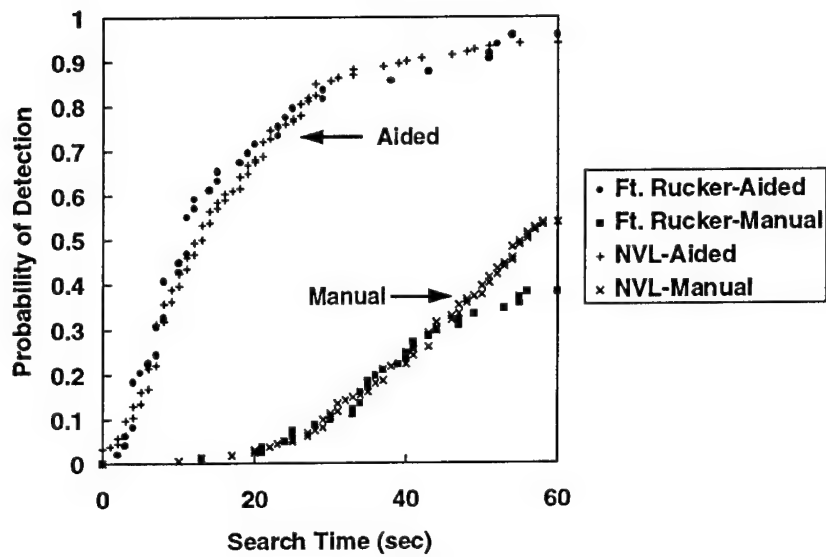


Figure 3. Probability of detection versus time for the Ft. Rucker and NVL observer sets. Both aided and manual modes shown.

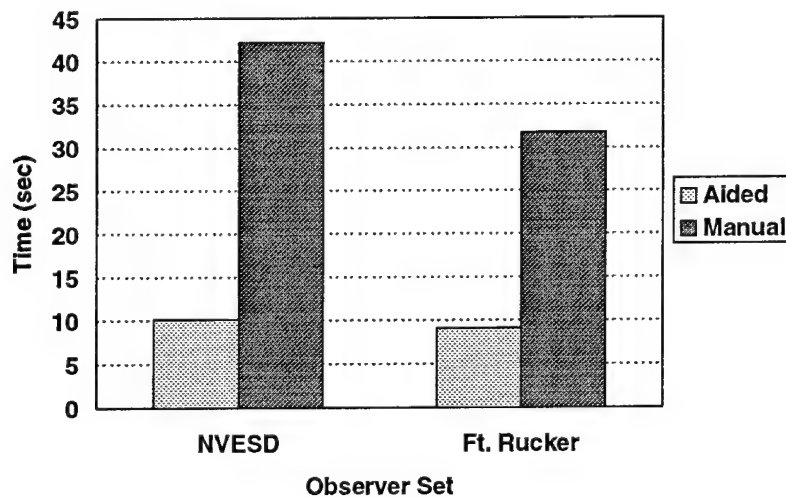


Figure 4. Time to first target detection for each observer set and each mode (aided and manual). Note significant decrease in time to first detection in aided mode.

Figure 5 shows the false alarm rate, i.e. false alarms vs. time (the number of false alarms has been normalized to the total FOR). It is interesting to note that the false alarm rate increases linearly with time, suggesting that human observers tend to find false alarms in a cluttered scene at a uniform rate. However, it is also observed that the aided rate is lower than the manual rate. Why this is so is not exactly clear as both sets of imagery were taken from identical backgrounds. It may be that the presence of ATR cues within the imagery helps to frame the observers' expectations of what a target should look like, or perhaps the observers just spent more time looking at the cues within the imagery and less time looking at the rest of the scene. Prior conclusions reached by Nicholl (1992) in analyzing the first NVESD search experiment (Blecha et al., 1991) suggest the latter mechanism; however, the very low cue density used here (less than one per four FOV's compared to a *minimum* of 5 per FOV in the Blecha experiment) means there weren't very many cues for the subjects to examine during the search. Furthermore, the subjects had presumably seen the cues previously in the chip display. It seems apparent that subjects must have spent substantial time looking at the background in the aided mode. A cued search experiment that incorporates eye tracking would resolve the issue of where the subjects are actually looking during search.

Analysis of the detailed results of the observer trials showed that the false alarm *rejection* ratio during chip review was approximately 12:1. That is, for every 12 ATR false alarms chips, the human observers reported one of the chips as a false alarm. This is substantially higher (better) than was measured in previous experiments, and suggests value added from the chip display in organizing the ATR results into an easily comprehensible list (and allowing ready comparison between real target chips and false target chips). One additional useful result obtained by analysis of the detailed timing data is that observers spent an average of 21 sec in the chip mode during aided search. This works out to about 6 seconds per chip, which includes both the cognitive process of "detecting" the target and the mechanical processes of registering the target as a "positive", changing its class, etc. Unfortunately, the method of recording the data used for the study does not permit separating the time allocation between the cognitive and mechanical tasks.

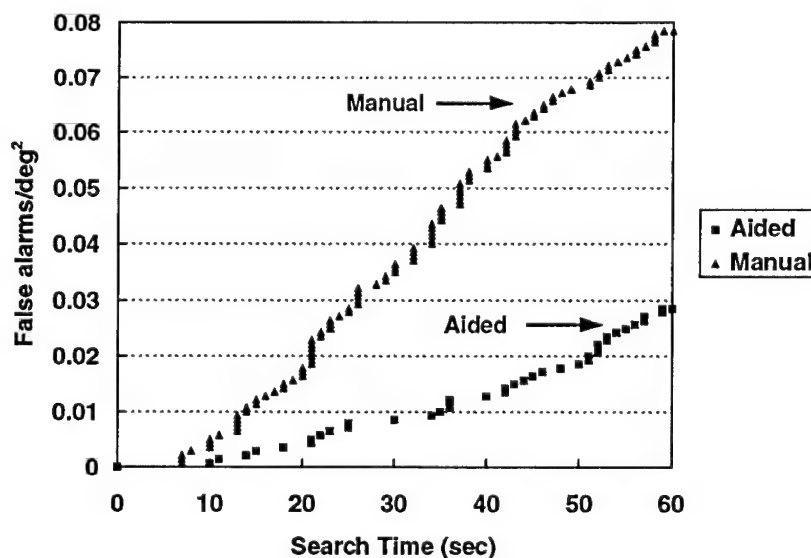


Figure 5. False alarms generated as a function of search time. Note that the slope of each curve is approximately linear.

4. Summary & Conclusions

Experiments were conducted to assess search performance in time compressed scenarios applicable to a reconnaissance/attack helicopter or a real-time UAV ground station. The target search performance of two sets of subjects was evaluated in two modes: a manual mode where target detection was solely dependent on the FLIR imagery and the observer, and an aided mode where the sensor imagery was supplemented with ATR cues. Digitally recorded 2nd Generation FLIR imagery and actual ATR results from the MSAT-Air ATD were used to construct the test set, while the man-machine interface was similar to that used on the actual MSAT-Air aircraft. The results were scored and analyzed for probability of detection vs. time, false alarm rate vs. time, and time to first target. The following conclusions can be drawn from the results:

1. *At the ATR performance levels evaluated, detection performance was significantly enhanced in the aided mode.* At the end of one minute, the subjects detected over 90% of the targets in the aided mode but less than 50% of the targets in the manual mode.
2. *False alarms were generated by the observers at a relatively constant rate (false alarms/sec); however, the rate was substantially lower in the aided mode than in the manual mode (even accounting for the false alarms from the ATR).* To understand this phenomenon, a follow-on test using an eye tracker is recommended.
3. *The overall time to detect the target was improved 3-4 fold in the aided mode.* This is in part due to the design of the man-machine interface, and can be critical in real time, high threat conditions.
4. *The value of ATR is magnified in short time searches.* The overall system concept and man-machine interface are essential to achieving good aided performance in time compressed scenarios.

References

1. Lloyd, E., Gercken, W., Jones, E., Gahler, M., Wright, R., Mayes, K. and Reago, D., *MSAT-Air Final Report*, Night Vision and Electronic Sensors Directorate, Ft. Belvoir, VA, April 1995.
2. Rodkey, Hostetler, Heinrich and Welby, *Automatic Target Recognition for Army Deep Strike Operations*, Proceedings of the Fourth Target Recognition Systems and Technology Symposium, Vol. II, pp. 21-42, March 1995.
3. Bohan, D., Haskett, H. and Shaffer, D., *Technical Assessment for Automatic Target Recognition*, Night Vision and Electronic Sensors Directorate, Ft. Belvoir, VA, April 1996.
4. Blecha, B., Do-Duc, H., O'Kane, B., and Crenshaw, M., *Aided Versus Unaided Search Experiments to Further Our Understanding*, Proceedings of the 1991 IRIS Passive Sensors Specialty Group, March, 1991.
5. O'Kane, B., Wolf, K., and Bonzo, D., *Aided Target Recognition/Cues as an Aid to Scouts*, Night Vision and Electronic Sensors Directorate, Ft. Belvoir, VA, March 1996.
6. Nicholl, J., *Performance in the Second NVEOD Aided Versus Unaided Search Experiment*, Paper P-2772, Institute for Defense Analysis, Alexandria, VA, December 1992.
7. Lawson, W., Cassidy, T. and Ratches, J., *A Search Prediction Model*, Proceedings of the IRIS Specialty Group on Imaging, June 1978.

HIGH-RESOLUTION ACOUSTIC DIRECTION-FINDING ALGORITHM TO DETECT AND TRACK GROUND VEHICLES

Tien Pham, Brian Sadler and Manfai Fong
US Army Research Laboratory
Adelphi, MD 20783-1197

Donald Messer
The American University
Washington, DC 20016-8101

Abstract

For our experiments, we use an incoherent wideband adaptive signal subspace algorithm for high-resolution aeroacoustic direction-finding. In this paper, we present experimental results for a circular array. In particular, the wideband incoherent MUSIC algorithm shows distinct improvements, as expected, over conventional beamforming in terms of accuracy and the ability to resolve multiple ground targets. Computational complexity is an issue, but can be reduced by using alternatives to full eigen-analysis for real-time implementation in the ARL sensor testbed.

1. Introduction

This paper describes work on array signal processing algorithms using small baseline acoustic arrays to passively detect, estimate direction of arrival (DOA), track, and classify moving targets for real-time implementation in the ARL sensor testbed [1]. Acoustic detection and tracking of ground vehicles in a battlefield environment is a challenging problem. Any dominant spectral lines in the acoustic signatures are nonstationary due to engine load and RPM changes during maneuvering. Terrain, atmospheric effects, and propagation characteristics that vary with time also produce significant signal variability. Figure 1 show the spectrogram of a maneuvering tracked vehicle during a test run conducted at Aberdeen Proving Ground (APG) in 1995. (Note the lack of energy beyond 150 Hz for this test run.) The usable frequency range for signal processing of ground vehicles is limited approximately to the range of 20 to 200 Hz for the detection range of interest (dominated below 20 Hz by wind noise, and dominated above 200 Hz by poor propagation characteristics). Sensor array baselines are physically constrained by system requirements and lack of spatial coherence to less than 10 ft. Because of the limited baseline and nonstationarities, conventional beamforming techniques provide poor spatial resolution, motivating the use of adaptive high-resolution DOA algorithms.

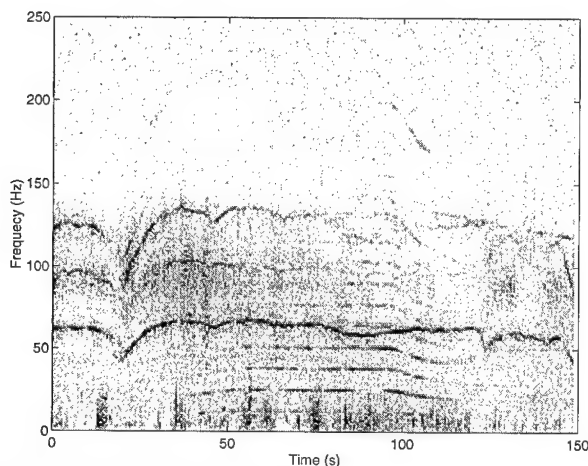


Figure 1. Spectrogram of an armored personnel carrier (APC).

Current efforts are focused on wideband signal subspace DOA algorithms such as incoherent wideband MUSIC [2, 3, 4]. Validation is accomplished using experimental data, and successful algorithms may then be implemented in the sensor testbed. Experimental results using a circular array of six sensors (plus one at the array center), with a diameter of 8 ft are presented for wideband MUSIC, illustrating its performance versus the delay-sum (DS) beamformer.

2. Sensor Testbed

The sensor testbed is a distributed platform used to prototype and evaluate new sensor technology in both hardware and software applications. Multiple-sensor arrays transmit detection and DOA information back to a central gateway. The gateway performs data fusion by collating DOA information with a tracker, and transmits results back to the combat information processor (CIP) for display on a graphical user interface. The testbed has been designed so any type of sensor or array geometry can be used, as long as it adheres to a predetermined communications protocol. The gateway is designed so that any data fusion algorithm can be used, as long as it uses predetermined communication protocols [1].

Currently, the testbed is being used to test acoustic-detection algorithms at the sensor-array level and an acoustic data-fusion algorithm at the gateway level. Four to six circular microphone arrays are placed, and vehicles are driven around a prescribed course inside the test site area. The CIP remotely displays how well the sensor arrays are detecting the vehicles, and how well the data fusion algorithm at the gateway is correlating the multiple-sensor arrays reports. A 9600-baud packet radio is used for the communications link between the sensor arrays and the gateway, and between the gateway and the CIP. Each microphone array has an embedded PC for determining the DOA. Figure 2 shows a typical CIP display for one update.

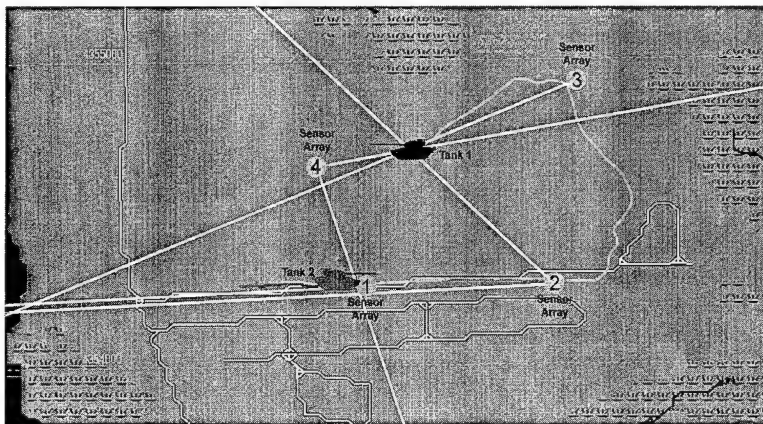


Figure 2. CIP display shows the sensor positions on a terrain map of the test area and the estimated DOA to the targets.

3. High-Resolution Direction-Finding Algorithm

3.1 Background

Conventional DS beamforming has been implemented in the acoustic testbed at the sensor-array level. The sensor array's PC collects acoustic data at a 2-kHz sampling rate from the microphones and estimates the DOA. The DOA estimates are updated and sent back to the gateway once a second for further processing. At the gateway, the DOA reports from multiple sensors are correlated to form a track for each vehicle. In addition, the gateway saves all DOA reports for post-test analysis. Vehicle tracks calculated by the gateway, along with DOA reports from the sensors, are transmitted back to the CIP for display and analysis by the system user (see figure 2). The DS beamformers perform well for a single-source environment; however, their ability to resolve closely spaced sources is limited by the small baselines of the arrays. Therefore, the resolution of the DS beamformer is greater than 5 degrees (based on the parameters used).

3.2. Implementation of MUSIC

Figure 3 shows the steps for implementing incoherent wideband MUSIC. To overcome the nonstationary nature of the acoustic source, the data is segmented before processing into fixed blocks, and stationarity is assumed over each data block. We have found that the assumption of signal stationarity is reasonable for intervals on the order of 1 second or

less. Over each processing interval, it is assumed that a single frequency bin is occupied by only a single source. This takes advantage of the nonstationarity of the sources, and simplifies the MUSIC algorithm. Although a source may be masked in a particular processing interval, this effect is unlikely to continue because the sources are changing frequencies as a function of time. Preprocessing is required to adaptively select the operating frequencies (the ω_k 's). Let $y_i(n)$ denote the output of the i th sensor from an array of N sensors, and let $Y_i(\omega)$ denote $DFT\{y_i(n)\}$. The average sum of $|Y_i(\omega)|^2$ is then obtained to adaptively select frequency bins of interest. This can be performed in a variety of ways, from simple thresholding based on frequency bin signal-to-noise ratio (SNR), to more complex schemes, such as harmonic association.

After the ω_k 's have been selected, the next step is formation of the estimated spatial correlation matrix for each ω_k , over each data block, given by

$$\hat{R}_Y(\omega_k) = \hat{Y}(\omega_k)^H \hat{Y}(\omega_k), \quad (1)$$

where $\hat{Y}(\omega_k) = [Y_1(\omega_k), Y_2(\omega_k), \dots, Y_N(\omega_k)]^T$ and H is the Hermitian operator (complex conjugate transpose). Once $\hat{R}_Y(\omega_k)$ is formed for each ω_k , the narrowband MUSIC algorithm is applied repeatedly [3]. The first step is eigenanalysis of $\hat{R}_Y(\omega_k)$ to obtain the noise subspace. Taking $\hat{R}_Y(\omega_k)$ to be $N \times N$ then, by the assumption that only one source can occupy a frequency bin over a processing interval, the noise subspace consists of the $N-1$ eigenvectors corresponding to the $N-1$ smallest eigenvalues of $\hat{R}_Y(\omega_k)$, and these form $\hat{U}_n(\omega_k)$. The MUSIC beampattern is then computed. For each look angle θ , the beampattern is given by

$$\hat{P}_{MUSIC}(\theta) = [E(\omega_k, \theta)^H \hat{U}_n(\omega_k) \hat{U}_n(\omega_k)^H E(\omega_k, \theta)]^{-1}. \quad (2)$$

The array manifold or steering vector, $E(\omega_k, \theta)$, is defined as

$$E(\omega_k, \theta) = [e^{2\pi j \Delta t_1}, e^{2\pi j \Delta t_2}, \dots, e^{2\pi j \Delta t_N}]^T, \quad (3)$$

where $\Delta t_i = \frac{d}{c} \sin \phi_i$, $\phi_i = \theta - \alpha_i$, where α_i is the relative angle to the normal for sensor $i = 1, 2, \dots, N$, d is the radius of the circular array, and c the speed of sound in air.

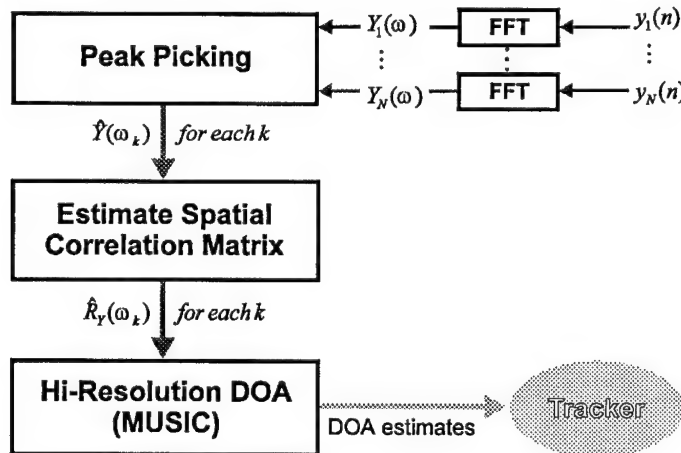


Figure 3. Implementation of the MUSIC algorithm.

After the beampatterns are calculated for all ω_k 's, they are incoherently averaged together to form a resulting MUSIC beampattern [2, 3]. Finally, the source DOA's are estimated by selecting the angles corresponding to the peaks in the resulting averaged beampattern. For each update, the raw DOA estimates from the sensor arrays are transmitted to the gateway for tracking. The tracker correlates the DOA estimates from the sensor arrays, and maintains a tracking

history for each acoustic source. Currently, the testbed uses an alpha-beta tracker [1], which is robust with respect to missing or outlying DOA estimates.

3.3. Computational Issues

The computational complexity of MUSIC is governed mainly by the eigen-analysis calculation, which is $O(N^3)$. However, this can be reduced by applying methods that calculate the signal subspace only, as opposed to the full eigen-analysis, because in our implementation the signal subspace is assumed to consist of one component only. By further assuming that only one source occupies a frequency bin, the algorithmic complexity and the estimation of the number of sources are simplified. The complexity can be reduced even further by precomputing $E(\omega_k, \theta)$ for all the frequency bins in the range of interest at an additional memory cost. An alternative method is to use a coarse beamformer (e.g., DS) during preprocessing to reduce the number of look angles. Therefore, $\hat{P}_{MUSIC}(\theta)$ is computed for fewer values of θ . In addition, a simple and computationally efficient form of $E(\omega_k, \theta)$ has been implemented. Other more complex forms of $E(\omega_k, \theta)$ can be used to incorporate the sensitivity and gain calibration of the sensors. However, the response of the sensors can change over time due to changes in the atmosphere, requiring the sensors to be recalibrated, and the parameters used in $E(\omega_k, \theta)$ to be updated and stored periodically (which is generally not feasible for low-cost unattended systems).

Other methods (e.g., harmonic line association (HLA)) can be used to limit the number of operating frequencies. The HLA algorithm detects the frequency bins that are above a set SNR threshold level. The bins are then sorted and combined into harmonic groups. Each harmonic group contains frequency bins that are harmonically related to a fundamental frequency of an engine rate. The HLA will declare a harmonic group valid (i.e., a source is detected) if there are at least m ($m = 2, 3, 4$ or 5) components in that group; otherwise, there is no source detected. The use of HLA in this way, however, requires that the acoustic sources belong to a group of known sources [1].

4. Data Analysis and Experimental Results

The raw DOA results for DS and wideband MUSIC algorithms are presented for ground vehicles traveling around a 2-km² area of open grass field. For each test run, one of the vehicles was equipped with a GPS sensor to provide accurate positioning ground truth. For comparison purposes, both algorithms are performed over $\theta \in [0, 360]$ degrees in 1-degree increments. The mean squared error (MSE) and the mean absolute error (MAE) for each test run are computed with the outliers removed [2]. Figure 4 shows raw DOA results for a 250-second 1-target test run. Note that the number of erroneous estimates is much less for MUSIC, which will improve detection and tracking of multiple sources. In general, over a number of test runs conducted during this field test at APG, $MSE \approx 20$ and $MAE \approx 3.5$ for DS, while $MSE \approx 3.5$ and $MAE \approx 1.5$ for MUSIC.

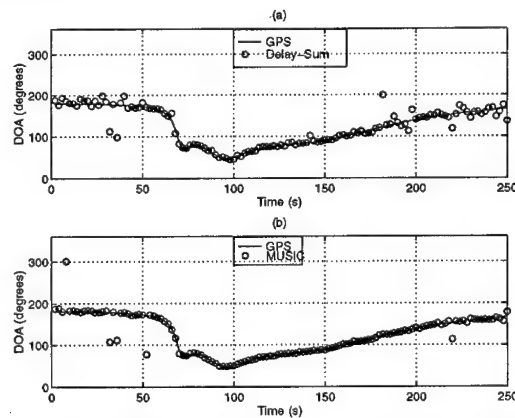


Figure 4. Raw DOA estimates versus GPS ground truth for (a) DS and (b) MUSIC beamformers.

Figure 5 shows the beampatterns of one processing interval (1 second) for DS and MUSIC for a 2-target test run. The sources are located at $\theta = 50$ and $\theta = 180$, respectively. Individual beampatterns were calculated for a set of 10 frequency bins, with the highest SNR and the resulting 10 beampatterns averaged together [2]. Note the sharpness of the peaks in the MUSIC beampattern in comparison to the DS beampattern, demonstrating the high resolution of the wideband MUSIC method for this problem. If the sources are closely located, the DS beamformer will not be able to detect and separate them.

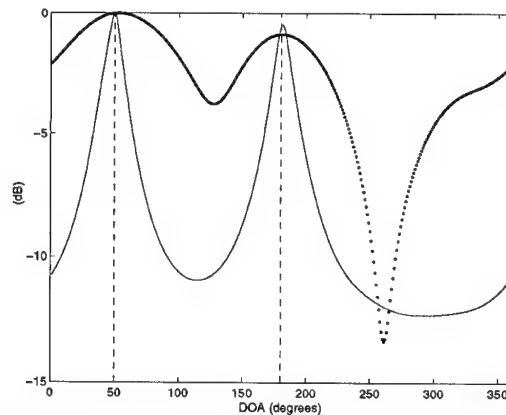


Figure 5. Beampatterns for two sources located at $\theta = 50$ and $\theta = 180$ for DS (dotted) and MUSIC.

Conclusions

Experimental results and analyses for both DS and MUSIC algorithms were presented. The acoustic sources are, generally, characterized as a sum of narrowband frequency components. Given adequate SNR, incoherent wideband MUSIC performed very well; produced more accurate DOA estimates; and yielded sharper distinct peaks in the beampattern, in comparison to DS. The disadvantage of implementing MUSIC is the higher computational cost. The computational complexity is governed by the eigen-analysis calculation, which is $O(N^3)$. Alternative methods were discussed to speed up the MUSIC algorithm for real-time implementation. Other computationally efficient signal subspace algorithms have been investigated prior to MUSIC (e.g., ESPRIT algorithm) [5]. However, ESPRIT generally requires twice the number of sensors as MUSIC, and places restrictions on array geometries. Current work includes high-resolution coherent wideband array processing to achieve higher processing gain during time intervals of low SNR [2, 4, 6]. Future work of interest includes reducing computation by exploiting the radial symmetry of a circular array [7], and effects of calibration and sensor placement errors [8].

References

- [1] N. Srouf and J. Robertson, "Remote Netted Acoustic Detection System: Final Report," ARL-TR-706, US Army Research Laboratory, Technical Report, Adelphi, MD (May 1995).
- [2] T. Pham and B. Sadler, "Adaptive wideband aeroacoustic array processing," *8th IEEE Statistical Signal and Array Processing Workshop (SSAP96)*, Corfu, Greece (24-26 June 1996).
- [3] M. Wax, T. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Trans. ASSAP*, Vol. 32, pp. 817-827, August 1994.
- [4] T. Pham and B. Sadler, "Aeroacoustic wideband array processing for detection and tracking of ground vehicles," *30th Meeting of the Acoustical Society of America*, St. Louis, MO (27 November - 1 December 1995).
- [5] T. Pham and B. Sadler, "Acoustic tracking of ground vehicles using ESPRIT," *SPIE Proceedings, Automatic Object Recognition V*, Vol. 2485, pp. 268-274, Orlando, FL (19-21 April 1995).
- [6] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources," *IEEE Trans. ASSP*, Vol. 33, pp. 823-831 (August 1985).
- [7] M. Doron, E. Doron, and H. Weiss, "Coherent wide-band processing for arbitrary array geometry," *IEEE Trans. SP*, Vol. 41, No. 1, pp. 414-417 (January 1993).
- [8] A. Swindlehurst and T. Kailath, "A performance based analysis of subspace-based methods in the presence of model errors, part 1: the MUSIC algorithm," *IEEE Trans. SP* Vol. 40, No. 7, pp. 1758-1774 (July 1992).

NONLINEAR AND QUANTUM OPTICS OF MULTICOMPONENT MEDIA

Michael E. Crenshaw* and Charles M. Bowden

Weapons Sciences Directorate, Missile Research, Development, and Engineering Center
U. S. Army Missile Command, Redstone Arsenal, AL 35898-5248

ABSTRACT

At densities that are typical of condensed matter, a propagating electromagnetic field mediates interactions between polarizable constituents of a material in what is known as the local-field effect. Previous investigations of the local-field effect have been limited to a single polarizable component of a nonlinear material. For multicomponent media, we have found that the interaction of laser radiation with an optically nonlinear component of condensed matter is fundamentally altered by the presence of another polarizable component as a result of the local-field effect. Novel effects that we have discovered include local-field enhancement effects, local cooperative decays, and coherence transfer and coherence exchange processes.

I. INTRODUCTION

Nonlinear and quantum optics are the underpinnings of modern laser technology. Essentially, these areas of optics are concerned with the interaction of intense laser radiation with matter. In addition to lasers, recent investigations have focused on phenomena that can form the basis for optical devices, such as the optical counterpart of the transistor, which would constitute the building blocks for all-optical-light-control-by-light digital optical data processing. All-optical digital systems offer the possibility of very high integrity systems, low noise, wide bandwidth, high speed, bulk processing, and elimination of optical-to-electronic interfaces. While much of the experimental and theoretical research in nonlinear and quantum optical device phenomenology has been performed in dilute vapors of atoms, fieldable devices that are compact, robust, and cost effective will have to be realized in condensed matter. At densities that are typical of condensed matter, an electromagnetic field mediates interactions between the polarizable particles of a material in what is known as the Lorentz local-field effect. To date, however, the Lorentz local-field condition (LLFC) has only been applied to a single polarizable component of an optically nonlinear material. Citing the examples of ions embedded in a crystal host and multiple bands in semiconductors, we note that optically nonlinear condensed matter often contains more than one polarizable component, even though only one is of primary interest. Here, we investigate the local-field effect for optically nonlinear materials which have more than one polarizable component. We find that the equations of motion for the interaction of laser radiation with an optically nonlinear component of condensed matter are fundamentally altered by the presence of another polarizable component. Novel effects that we have discovered include local-field enhancement effects, local cooperative decays, and coherence exchange assisted and impeded processes.

II. THE LORENTZ LOCAL-FIELD CONDITION

An optical material can be regarded as being composed of a large number of polarizable particles (atoms, molecules, excitons, ions, etc.) embedded in the vacuum. Then the field acting on a particle, the microscopic Lorentz local field, must be distinguished from the macroscopic field that is obtained by performing an average of the microscopic field over a region of space that contains a great number of particles. The macroscopic field, the field that appears in Maxwell's equations, contains all fields present, while the Lorentz local field does not contain the self-field of the particle under consideration. The Lorentz local-field condition (LLFC)

$$\mathcal{E}_L = \mathcal{E} + \left(\frac{4\pi}{3} + s \right) \mathcal{P} \quad (1)$$

relates the microscopic local field \mathcal{E}_L to the macroscopic field \mathcal{E} and polarization \mathcal{P} , where s is a structure factor to account for non-cubic crystal symmetry. This relation has been shown to apply to both stationary and propagating fields in nonlinear media (Bowden and Dowling, 1993), as well as in linear media, and it is generally assumed that this relation, with $s = 0$, is valid for isotropic homogeneous media. For linearly polarizable particles, the LLFC leads to the familiar Clausius-Mossotti-Lorentz-Lorenz relation that associates a macroscopic dielectric function with the microscopic polarizability and the number density of particles. In

the Drude model, in which atoms are modeled as harmonic oscillators, the LLFC leads to a renormalization of the resonance frequency that is associated with the introduction of the plasma frequency into the macroscopic equation of motion. For atoms (or ions, molecules, or other particles) that are modeled quantum mechanically as consisting of discrete energy levels connected by electric dipole transitions, interactions between atoms, known as near dipole-dipole (NDD) interactions, occur within the context of the LLFC when the density of atoms is sufficiently large. The NDD interactions give rise to a wide array of novel and interesting phenomena in nonlinear and quantum optics including intrinsic optical bistability, quasiadiabatic following and inversion, and piezophotonic and magnetophotonic switching.

For media composed of more than one component, the polarization that appears in Eq. (1) is the sum of the partial polarizations of each component. Thus, the interaction of one component with the propagating Maxwell field is coupled to the dynamics of other components through the macroscopic polarization. To date, however, applications of the local field condition have been limited to a single polarization component of an optically nonlinear material.

III. TWO-LEVEL ATOMS EMBEDDED IN A LINEARLY POLARIZABLE HOST

Recent experiments by Hehlen, Güdel, Shu, Rai, Rai, and Rand (HGS³) (1994) have provided dramatic verification of theoretical predictions (Hopf, Bowden, and Louisell, 1984) of intrinsic optical bistability due to NDD interaction. These experiments were performed with Yb³⁺ ions in a Cs₃Y₂Br₉ crystal at densities sufficient to cause strong local-field effects. While the HGS³ experiments were performed with resonant systems embedded in a crystal host, previous theoretical treatments of NDD interactions have assumed a dense collection of resonant systems *in vacuo*.

Here, as in Crenshaw and Bowden (1996), we consider the case in which a dense collection of two-level atoms is embedded in a linearly polarizable host and derive equations of motion for the interaction of laser radiation with the material. The modified optical Bloch equations that are obtained will have almost universal applicability for dense media because, in order to achieve the required densities for NDD interaction effects, resonant atoms will almost invariably be embedded in a solid-state host, as in the HGS³ experiments. We find enhancement of the NDD interaction due to the presence of the host dielectric, which is especially significant due to the fact that NDD effects are only important at sufficiently high densities and large oscillator strengths, and the appearance of additional coherence terms which are quite novel in that they represent the *nonlinear* interaction of dipoles mediated by the absorptive component of the dielectric function of the host.

The dynamics of the two-level systems are described by the generalized Bloch equations in the rotating-wave approximation (Bowden and Dowling, 1993):

$$\frac{\partial R}{\partial t} = i\Delta R - \frac{i\mu}{2\hbar} \left(\mathcal{E} + \frac{4\pi}{3}\mathcal{P} \right) W - \gamma_{\perp} R \quad (2a)$$

$$\frac{\partial W}{\partial t} = -\frac{i\mu}{\hbar} \left[\left(\mathcal{E}^* + \frac{4\pi}{3}\mathcal{P}^* \right) R - \left(\mathcal{E} + \frac{4\pi}{3}\mathcal{P} \right) R^* \right] - \gamma_{\parallel}(W - W_{eq}). \quad (2b)$$

The macroscopic, spatially averaged, atomic variables in the rotating frame of reference are $R = \langle \rho_{21} e^{i\omega t} \rangle_{sp}$, $R^* = \langle \rho_{12} e^{-i\omega t} \rangle_{sp}$, and $W = \langle \rho_{22} \rangle_{sp} - \langle \rho_{11} \rangle_{sp}$. Here, $\langle \dots \rangle_{sp}$ corresponds to a spatial average over a volume of the order of a resonance wavelength cubed and the ρ_{ij} are the density matrix elements for a two-level system with a lower state $|1\rangle$ and an upper state $|2\rangle$. Also, μ is the transition dipole moment, $\Delta = \omega - \omega_0$ is the detuning from resonance, γ_{\perp} is the dipole dephasing rate, γ_{\parallel} is the population relaxation rate, and W_{eq} is the population difference at equilibrium.

The polarization of the medium, which is composed of an isotropic homogeneous distribution of resonant systems in a linearly polarizable host, is $\mathcal{P} = \mathcal{P}^{bg} + \mathcal{P}^{res} = \alpha N_a \mathcal{E}_L + \mathcal{P}^{res}$, the sum of a background polarization, linear in the local field, that is due to the host material and a nonlinear polarization that is due to the resonant systems. Applying the LLFC, one obtains

$$\mathcal{P} = \frac{\varepsilon - 1}{4\pi} \mathcal{E} + \frac{\varepsilon + 2}{3} \mathcal{P}^{res}, \quad (3)$$

where ε is the complex dielectric function of the host and the structure factor has been assumed to be negligible for the mixture of homogeneous distributions of particles.

Then, using Eq. (3) to eliminate the polarization, the generalized macroscopic Bloch equations become

$$\frac{\partial R}{\partial t} = i(\Delta - \ell \epsilon W)R - \frac{i\mu}{2\hbar} \ell \mathcal{E} W - \gamma_{\perp} R \quad (4a)$$

$$\frac{\partial W}{\partial t} = -\frac{i\mu}{\hbar} (\ell^* \mathcal{E}^* R - \ell \mathcal{E} R^*) - 2i(\ell^* - \ell) \epsilon |R|^2 - \gamma_{\parallel} (W - W_{eq}), \quad (4b)$$

where $\mathcal{P}^{\text{res}} = 2N\mu R$ is the nonlinear polarization, $\epsilon = 4\pi N\mu^2/3\hbar$ is the NDD interaction parameter, N is the number density of two-level systems, and $\ell = (\epsilon + 2)/3$ is the local-field enhancement factor arising from the elimination of the total polarization using Eq. (3).

Equations (4) display several local-field effects: *i*) There is an apparent enhancement of the magnitude of the field that drives the atoms that is largely negated by Ewald-Oseen extinction. The overall phase shift of the field has no effect on the dynamics. *ii*) The inversion-dependent detuning (nonlinear Lorentz frequency shift) that is due to the NDD interaction is intensified by the real part of the local-field enhancement factor ℓ . *iii*) Cooperative decay terms appear due to the interaction of near dipoles mediated by the imaginary component of the dielectric function of the host material. It should be noted that the modification of the optical Bloch equations to include local-field effects requires a re-interpretation of the vector model for a two-level atom in terms of an averaged, macroscopic Bloch vector in which the torque vector contains reaction from the Bloch vector as a manifestation of quantum coherences.

Because NDD effects are only important at sufficiently high densities and large oscillator strengths, it is significant that the local-field enhancement factor increases the effect of the inversion-dependent detuning. For example, Friedberg, Hartmann, and Manassah (1989) derived a threshold condition for intrinsic optical bistability in a dense vapor, namely $\epsilon > 4\gamma_{\perp}$. This density threshold condition cannot be controlled in a vapor of two-level systems because the dephasing rate increases in direct proportion to the density due to collisional broadening. For two-level systems in a dispersionless dielectric, the threshold condition $\ell\epsilon > 4\gamma_{\perp}$ can be satisfied by a smaller value of ϵ , because *i*) the inversion-dependent detuning is enhanced and *ii*) in condensed matter, the homogeneous linewidth is no longer restricted to the formula for a collisionally broadened vapor. Significantly, it is in a similar case that intrinsic optical bistability, analyzable by a two-level model, was observed experimentally (Hehlen, *et al.*, 1994).

A linear dielectric which is dispersive must also be absorptive in accordance with the Kramers-Kronig relations. Then the dielectric function, local-field enhancement factor, and index of refraction are complex quantities. The imaginary part of the enhancement factor appears as a coefficient, along with the NDD parameter, of bilinear products of the macroscopic atomic variables in two terms of Eqs. (4). These terms correspond to local cooperative decay effects representing the interaction of near dipoles mediated by the imaginary component of the dielectric function of the host medium. Because the effects of NDD interactions can be manifested in films that are significantly thinner than a vacuum wavelength, the detrimental effects, absorption and heating, associated with the imaginary part of the index of refraction can be mitigated for dense media. Then, for a thin film of a strongly dispersive dielectric containing a dense collection of two-level atoms, one can expect local cooperative decay effects to play a significant role in the dynamics.

IV. THREE-LEVEL ATOMS IN A LINEARLY POLARIZABLE HOST

Recently, three-level systems have garnered considerable interest in the quantum optics community, due primarily to the theoretical prediction (Kocharovskaya, 1992) and experimental verification (Padmabandu, *et al.*, 1996) of lasing-without-inversion (LWI) in three-level Λ and V systems. It has been predicted that inversionless gain can be increased by up to 2 orders of magnitude if the density of three-level atoms is sufficiently large (Manka, *et al.*, 1994). Several processes, many with novel device applications, that occur only in dense collections of three-level atoms have been discovered (Wang, *et al.*, 1996). While to achieve the required densities, the three-level systems will almost invariably be embedded in a host material, the effect of the host has been ignored previously. We have investigated the case in which a dense collection of three-level atoms is embedded in a linearly polarizable host. Applying the LLFC, we find several local-field effects. *i*) There is an apparent enhancement of the magnitude and an overall phase shift of the field driving the atoms. *ii*) The coupling of the two polarization channels is intensified by the real part of the local-field enhancement factor ℓ . *iii*) Intrinsic cross-channel cooperative decays, as well as intrinsic cooperative decays, appear due to the interaction of dipoles mediated by the imaginary component of the dielectric function of the host. Because there are six equations for each of the ladder, V, and Λ configurations, the complete details of our analysis must be published elsewhere.

V. SEMICONDUCTORS

The recent widespread experimental capability to generate intense subpicosecond laser pulses coupled with the goal of cost effective optical devices that can operate at high bit rates have stimulated substantial interest in laser field-induced many-body collective and coherence effects in semiconductors. A set of Bloch-like equations of motion for semiconductor crystals, which include coherent Coulomb interactions and the density-of-states distribution to first order, can be reduced from the Hartree-Fock equations of a many-body electron-hole model of a two-band semiconductor. Although the semiconductor is modeled as consisting of an ensemble of two-level systems with a range of transition frequencies, semiconductor crystals typically exhibit a very large linear index of refraction as well. The linear index can be viewed as arising from the adiabatic elimination of the dynamics of additional bands, rather than as an embedding in a linearly polarizable host. Applying the LLFC to the reduced semiconductor Bloch equations (Bowden and Agrawal, 1995), we obtain

$$\frac{\partial R}{\partial t} = i(\Delta + (\epsilon_d - \ell\epsilon)W)R - \frac{i\mu}{2\hbar}\ell\mathcal{E}W - \gamma_{\perp}R - \mathcal{S}^2S \quad (5a)$$

$$\frac{\partial W}{\partial t} = -\frac{i\mu}{\hbar}(\ell^*\mathcal{E}^*R - \ell\mathcal{E}R^*) - 2i(\ell^* - \ell)\epsilon|R|^2 - \gamma_{\parallel}(W - W_{eq}), \quad (5b)$$

$$\frac{\partial S}{\partial t} = i(\Delta + \epsilon_d W)S + \ell\epsilon RU + \frac{\mu}{2\hbar}\ell\mathcal{E}U - \gamma_{\perp}S \quad (5c)$$

$$\frac{\partial U}{\partial t} = -\frac{\mu}{\hbar}[\ell^*\mathcal{E}^*S + \ell\mathcal{E}S^*] - 2\epsilon(\ell^*R^*S - \ell RS^*) - \gamma_c(U - \bar{U}), \quad (5d)$$

where ϵ_d is the Debye factor arising from the coherent Coulomb exchange interaction and \mathcal{S} is an expansion parameter. In Eq. (5a), we find an enhancement of the local-field correction to the carrier density-dependent renormalization of the band gap. As in the case of embedded two-level systems, Eqs. (5a) and (5b) exhibit local cooperative decays if the local-field enhancement factor is complex. The imaginary part, as well as the enhancement, of the NDD terms in Eqs. (5c) and (5d) have important implications for instabilities and chaos in semiconductor lasers (Bowden, Singh, and Agrawal, 1995).

VI. THE GENERALIZED BLOCH-DRUDE MODEL

To this point we have assumed that a constitutive relationship exists for the interaction of the host with the field, allowing us to obtain representative results without the necessity of specifying a particular model for the host, which is of secondary importance. Let us now assume that the host can be modeled as a collection of harmonic oscillators. Applying the LLFC, we obtain generalized Bloch-Drude equations

$$\frac{\partial R}{\partial t} = i(\Delta - \epsilon W)R - \frac{i}{2}\Omega W - \frac{i}{2}\frac{\omega_p^2}{3}\bar{X}W - \gamma_{\perp}R \quad (6a)$$

$$\frac{\partial W}{\partial t} = -i(\Omega^*R - \Omega R^*) - i\frac{\omega_p^2}{3}(\bar{X}^*R - \bar{X}R^*) - \gamma_{\parallel}(W - W_{eq}) \quad (6b)$$

$$\frac{\partial^2 \bar{X}}{\partial t^2} + (\gamma - 2i\omega)\frac{\partial \bar{X}}{\partial t} + (\omega_b^2 - \omega^2 - \omega_p^2/3 - i\omega\gamma)\bar{X} = (\Omega + 2\epsilon R), \quad (6c)$$

where the plasma frequency ω_p is defined by the relation $\omega_p^2 = 4\pi N_b e^2/m$, N_b is the number density of oscillators, ω_b is the resonance frequency of the oscillators, and $\bar{X} = m\mu X/e\hbar$ is a normalized coordinate.

Under certain conditions, Eq. (6c) can be adiabatically eliminated to yield

$$\bar{X} = \frac{\Omega + 2\epsilon R}{\omega_b^2 - \omega^2 - \omega_p^2/3 - i\omega\gamma}, \quad (7)$$

which is the usual Drude result for a dense collection of oscillators, but with a dynamic field renormalization arising from coherence transfer due to the LLFC. Inserting Eq. (7) into Eqs. (6a) and (6b), the equations of motion for a dense collection of two-level atoms embedded in a linearly polarizable host are obtained, where $\ell = 1 + (\omega_p^2/3)/(\omega_b^2 - \omega^2 - \omega_p^2/3 - i\omega\gamma)$. Now, ℓ has explicit physical meaning in the context of a particular physical model of the host polarizability.

VII. THE MULTICOMPONENT GENERALIZED BLOCH MODEL: COHERENCE EXCHANGE

Finally, let us consider the case of two species, a and b , of two-level atoms. The multicomponent generalized Bloch equations

$$\frac{\partial R}{\partial t} = i\Delta_a R - \frac{i\mu_a}{2\hbar} \left(\mathcal{E} + \frac{8\pi}{3} N_a \mu_a R + \frac{8\pi}{3} N_b \mu_b S \right) W - \gamma_{\perp a} R \quad (8a)$$

$$\frac{\partial W}{\partial t} = -\frac{i\mu_a}{\hbar} \left[\left(\mathcal{E}^* + \frac{8\pi}{3} N_a \mu_a R^* + \frac{8\pi}{3} N_b \mu_b S^* \right) R - \text{c.c.} \right] - \gamma_{\parallel a} (W - W_{eq}) \quad (8b)$$

$$\frac{\partial S}{\partial t} = i\Delta_b S - \frac{i\mu_b}{2\hbar} \left(\mathcal{E} + \frac{8\pi}{3} N_a \mu_a R + \frac{8\pi}{3} N_b \mu_b S \right) Z - \gamma_{\perp b} S \quad (8c)$$

$$\frac{\partial Z}{\partial t} = -\frac{i\mu_b}{\hbar} \left[\left(\mathcal{E}^* + \frac{8\pi}{3} N_a \mu_a R^* + \frac{8\pi}{3} N_b \mu_b S^* \right) S - \text{c.c.} \right] - \gamma_{\parallel b} (Z - Z_{eq}) \quad (8d)$$

contain terms that can be interpreted as a source of coherence exchange. Coherence exchange assisted and coherence exchange impeded processes, such as induced intrinsic optical bistability, can occur in a dense mixture of two species of two-level atoms. An even richer variety of coherence exchange processes are possible in multicomponent multilevel systems. More detailed discussion of coherence exchange processes must appear elsewhere.

For now, let us consider the case where Δ_b is sufficiently large compared to $\mu_b(\mathcal{E} + 8\pi N_a \mu_a R/3)\gamma_{\perp b}/\gamma_{\parallel b}$ that $Z \rightarrow -1$. This is the weak field or harmonic oscillator limit of the two-level atom. The equations of motion for species b collapse to

$$-i\frac{\partial S}{\partial t} + (-\Delta_b - \epsilon_b - \gamma_{\perp b})S = \frac{\mu_b}{2\hbar} \left(\mathcal{E} + \frac{8\pi}{3} N_a \mu_a R \right). \quad (9)$$

Under certain conditions, Eq. (9) can be adiabatically eliminated and Eqs. (8) reduce to the equations of motion for a dense collection of two-level atoms embedded in a linearly polarizable host, where, as in the generalized Bloch-Drude model, ℓ has explicit physical meaning in the context of a particular physical model of the host polarizability.

VIII. SUMMARY

We have described the derivation of self-consistent equations of motion for dense multicomponent media. We found that novel quantum coherence effects and coherence transfer effects occur in dense multicomponent materials as the polarizable components are coupled in the context of the Lorentz local-field condition. Detailed treatment of these effects appear in our recent publication (Crenshaw and Bowden, 1996) or will be treated in future publications.

REFERENCES

- Bowden, C. M. and G. P. Agrawal, *Phys. Rev. A* **51**, 4132 (1995).
- Bowden, C. M. and J. P. Dowling, *Phys. Rev. A* **47**, 1247 (1993).
- Bowden, C. M., S. Singh, and G. P. Agrawal, *J. Mod. Opt.* **42**, 101 (1995).
- Crenshaw, M. E. and C. M. Bowden, *Phys. Rev. A* **53**, 1139 (1996).
- Friedberg, R., S. R. Hartmann, J. T. Manassah, *Phys. Rev. A* **40**, 2446 (1989).
- Hehlen, M. P., H. U. Güdel, Q. Shu, J. Rai, S. Rai, and S. C. Rand, *Phys. Rev. Lett.* **73**, 1103 (1994).
- Hopf, F. A., C. M. Bowden, W. Louisell, *Phys. Rev. A* **29**, 2591 (1984).
- Kocharovskaya, O. *Phys. Rep.* **219**, 175 (1992).
- Manka, A. S., J. P. Dowling, C. M. Bowden, and M. Fleishhauer, *Phys. Rev. Lett.* **73**, 1789 (1994).
- Padmabandu, G. G., *et al.*, *Phys. Rev. Lett.* **76**, 2053 (1996).
- Wang, N., H. Rabitz, A. S. Manka, and C. M. Bowden, *Phys. Rev. A*, to appear (1996).

Mathematical and Computer Sciences

Parallel Finite Element Computation of Missile Flow Fields

W. Sturek* and S. Ray
U.S. Army Research Laboratory
Aberdeen Proving Ground, Maryland, 21005-5066

S. Aliabadi, C. Waters and T. Tezduyar
Aerospace Engineering and Mechanics
Army High Performance Computing Research Center
University of Minnesota
Minneapolis, Minnesota, 55415

Abstract

The compressible, Navier-Stokes flow solver developed by the authors at the Army HPC Research Center is used to study missile aerodynamics at supersonic speeds and high angles of attack for large Reynolds numbers flow conditions. The goal of this study is the evaluation of this Navier-Stokes computational technique for the prediction of separated flow fields around high length-to-diameter (L/D) bodies at moderate angles of attack and supersonic velocities. In particular, this paper addresses two issues: turbulence modeling within the finite-element computational technique; and efficient performance of the computational technique on two multiprocessor mainframes, the Thinking Machines CM-5 and CRAY T3D. Comparisons are shown between the computation and experiment for the supersonic ramp flow to evaluate the suitability of the turbulence model. Also, results of the computations for the missile flow fields are shown for laminar and turbulent viscous effects.

1. Introduction

The compressible Navier-Stokes flow solver (Aliabadi and Tezduyar (1995), Tezduyar et al. (1995), Tezduyar et al. (1994)) developed by the authors at the Army HPC Research Center (AHPCRC) is used to study missile aerodynamics at supersonic speeds and high angle of attack for large Reynolds numbers flow conditions. The goal of this study is the evaluation of this Navier-Stokes computational technique for the prediction of separated flow fields around high length-to-diameter (L/D) bodies at moderate angles of attack and supersonic velocities. In particular, this paper addresses two issues: turbulence modeling within the finite element computational technique; and efficient performance of the computational technique on two multiprocessor mainframes, the Thinking Machines CM-5 and CRAY T3D.

This effort is also participating in a joint collaborative working group (KTA 2-12) under the auspices of The Technical Cooperation Program (TTCP). TTCP is a joint program with the United States, United Kingdom, Canada, Australia, and New Zealand which was established for the purpose of exchange of technical information between government agencies of these five countries.

Since Army missiles and projectiles fly under conditions of moderate to high angles of attack at transonic and supersonic velocities, turbulent viscous effects are an important parameter which must be modeled accurately in order for the predicted results to be of value in the design process. Also, due to the complex geometry of missiles which includes fin surfaces for guidance and flight stability, significant computational resources in terms of memory and CPU time are required to achieve solutions. Thus, the efficient performance of the Navier-Stokes solver on large, scalable computing mainframes is also of keen interest.

Through participation in TTCP working group (KTA 2-12), a set of experimental data has become available for comparison to the computations. These comparisons have provided guidance for evaluation of the turbulent viscous effects. Initial comparisons indicate that laminar viscous computations are unable to adequately predict important features of the flowfield which include surface pressure distributions and the separated vortex flowfield. The KTA study has not

yet been completed so direct comparisons of the computational results with experiment will not be shown here. However, experimental measurements of a turbulent boundary layer in a supersonic compression ramp flow are available for evaluation of the ability of the computational technique to predict this turbulent viscous flow field.

2. Parallel Computation

The stabilized finite element formulation of the compressible Navier-Stokes equations is implemented on the distributed memory CM-5 and T3D supercomputers. The implementation on the CM-5 is based on the data-parallel computing paradigm and the CMF (Connection Machine Fortran) language. Our implementation on the T3D takes advantage of Parallel Virtual Machine (PVM) library.

For efficient computation, a mesh partitioning method is used on both supercomputers to minimize the interprocessor communication. In this method, the finite element mesh is partitioned into desired number of subdomains such that the number of the nodes at the interface of the subdomains is minimized. Then, the gather and scatter operations which are basically the transfer of data from the node-level to the element-level and vice-versa, are performed in two steps: on-processor and off-processor. In the on-processor step, the gather and scatter operations are local to the processor with no interprocessor communication. The communication between the processors which are now mostly related to nodes on the boundary of each subdomain is performed in off-processor step. On the T3D, the two-step gather and scatter operations are carried out using the PVM-based routines. On the CM-5, there are routines supported by the Connection Machine Scientific Software Library (CMSSL) which allow the user to perform two-step gather and scatter operations. This feature makes the finite element programming relatively easier on the CM-5 compared to the T3D.

The compressible Navier-Stokes finite element formulation gives rise to very large systems of coupled nonlinear equations which require the use of iterative strategies with update techniques such as GMRES for their solution. To further reduce the memory requirements, we use matrix-free iterations and thus eliminate the need to store element-level matrices (Aliabadi and Tezduyar (1995)). The overall performance of our codes is comparable on each supercomputer. The computational speed, measured for problems reported in this article is around 20 MFLOPS per processor on both machines.

3. Computational Results

3.1 Compression Ramp

The experimental study was performed at the Supersonic Wind Tunnels of the Army Ballistic Research Laboratory and is reported by Sturek (1970). The wind tunnel was a boundary layer channel with a cross section and height in the test section of 6x6 - inches. The free stream conditions are Mach = 3.5, total pressure = 53.35 psi, and total temperature = 540R which gives a Reynolds number for the free stream flow of 6.4 million per foot. Extensive, detailed measurements of the turbulent boundary layer velocity profiles, surface pressure, and skin friction are available for comparison with the computational results.

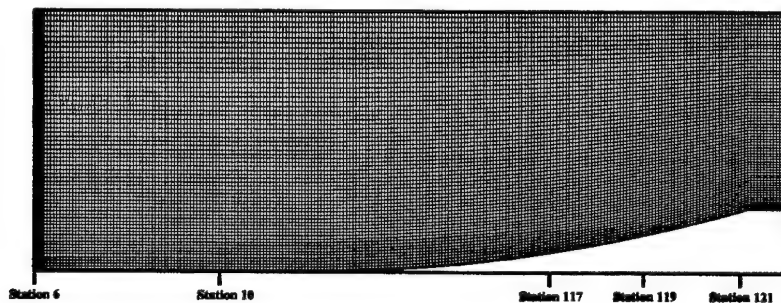


Figure 1. Compression ramp: finite element mesh (every other node plotted).

The computations were carried out using experimental data as the upstream boundary condition and assuming an outflow outer boundary. The computational grid is shown in figure 1 where only every other grid point is displayed for

clarity. Also shown in figure 1 are the identification and position of the test stations. Only comparison between the computations and experiment at the stations 10 and 119 are presented in this article. For more detail, please refer to Sturek (1996). A zero pressure gradient condition downstream of the initial test station exists for 5-inches, followed by the adverse pressure gradient compression ramp flow for 10-inches.

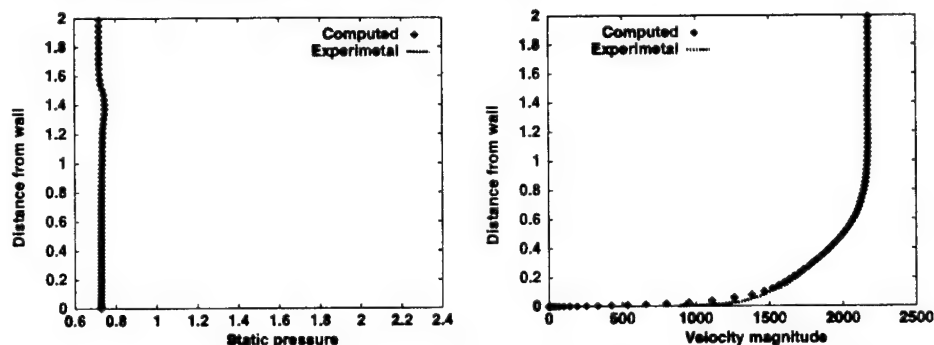


Figure 2. Compression ramp: comparison between the computed and experimental results at station 10.

Comparisons of the computations with experimental data for the boundary layer profile parameters of static pressure (psi) and velocity magnitude (fps) are shown in figures 2 and 3. The results indicate very good agreement at Station 10, the last station in the zero pressure gradient region. As the flow develops over the ramp, the agreement between computation and experiment, although quite good at Stations 117 and 119, is not acceptable at the last station on the ramp, Station 121. Although this could be an indication of inaccuracy for the turbulence model, it could also be the result of upstream influence resulting from the out flow boundary condition. Future studies will address this.

These results were achieved on the T3D using 16 processors and a finite element mesh with 500x150 elements. The number of equations solved at each time step is 381,600. For this problem, a converged solution was obtained after 400 pseudo-time steps which required 50 seconds of CPU time for each pseudo-time step.

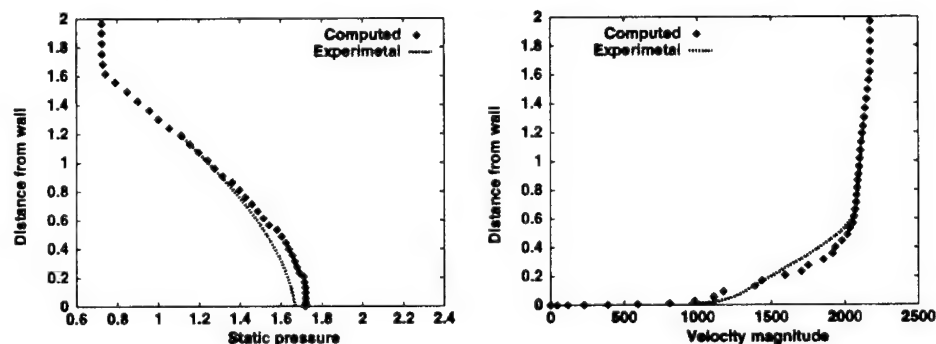


Figure 3. Compression ramp: comparison between the computed and experimental results at station 119.

3.2 Missile

As previously discussed, the missile test case is part of a collaborative study which is being carried out under the auspices of The Technical Cooperation Program Panel W2. The study includes the flow about a missile configuration at transonic and supersonic velocities and high angles of attack. Since this study has not been completed, comparisons with the experimental data will not be shown at this time. Instead, we will show comparisons between results for laminar and turbulent viscous effects and discuss the performance of the computational technique on the multiprocessor computers.

In this paper, preliminary results will be shown for free stream conditions of Mach 2.5 and 14-degrees angle of attack for laminar and turbulent viscous effects. figure 4 shows the model configuration along with contours of the flow field static pressure on the model surface at three axial stations in the cross-plane. The regions of high pressure on the

windward side at the nose and low pressure separated flow on the lee-side of the cylindrical portion of the model are clearly distinguishable.

Examples of the surface pressure at a series of axial stations are shown in figure 5. As the flow proceeds down the missile body, increasing regions of separated flow are seen. The laminar viscous results indicate more oscillations in the surface pressure on the lee-side of the model than the turbulent viscous results. These oscillations are a result of the effects of boundary-layer separation for which the laminar viscous layer is sensitive. At X/D greater than 6.5, separation prior to the 90-degree circumferential position is indicated for both the laminar and turbulent viscous results. These separated flow regions are very challenging for computational prediction and require careful evaluation of the computational results with the experimental measurements to assess the validity of the turbulence modeling. The results shown here are considered to be preliminary. Evaluations are ongoing and include study of the ability to predict separated outer flow field as well.

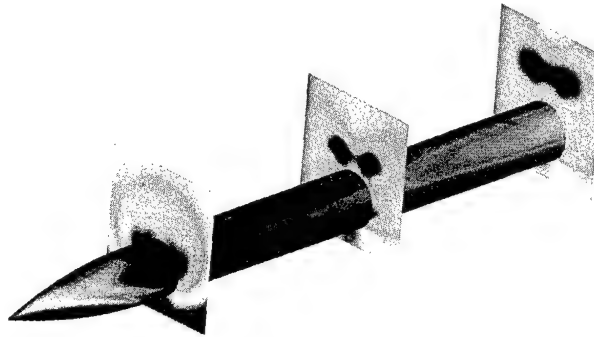


Figure 4. Missile: static pressure distribution on the missile surface and on three crossplane axial stations.

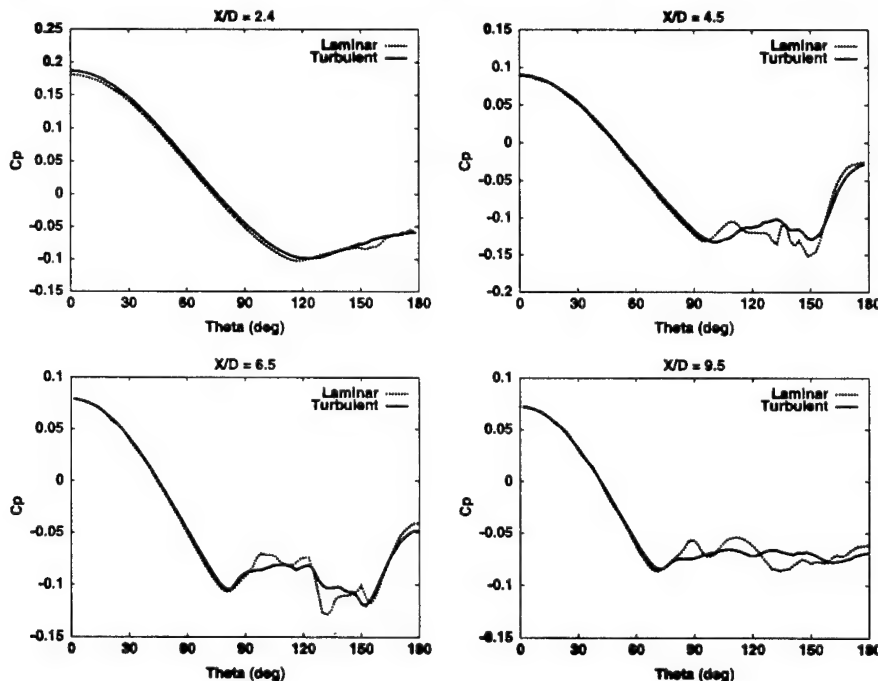


Figure 5. Missile: comparison of computed results on the T3D for laminar and turbulent viscous effects at various stations.

The computations for the missile configuration for both laminar and turbulent flows were accomplished on the T3D computer using 256 processors with a finite element mesh consisting of 944,366 nodes and 918,000 elements. This

resulted in the solution of 4,610,378 coupled, nonlinear equations at each pseudo-time step. Each pseudo-time step required 130 seconds of CPU time. The solutions were obtained in 400 pseudo-time steps.

Computations for the laminar and turbulent flows were also carried out on the CM-5 with 512 processors. The mesh used in these computations consists of 763,323 nodes and 729,600 elements. During each pseudo-time step, 3,610,964 coupled, nonlinear equations were solved. Each time step required approximately 50 seconds of CPU time and convergence was achieved after 150 pseudo-time steps. In this case, the laminar solution appears to be more turbulent-like than those obtained for the finer mesh CRAY T3D results shown in figure 5, and the turbulent results are only moderately different from the laminar results. This is further indication that additional effort is needed to evaluate the suitability and accuracy of the turbulence model used.

4. Conclusion

A computational study has been conducted to evaluate the ability of a finite element compressible, Navier-Stokes computational technique to perform computations for high Reynolds number flows of interest to Army missile and projectile flow fields. The results for predicting turbulent viscous effects are encouraging; however, additional evaluation is required to reach a meaningful assessment of the capability.

The computational performance achieved on the CM-5 and CRAY T3D computers indicate that highly efficient scalable performance has been achieved. This is an important consideration since missile configurations currently under development include features such as fins for stability and guidance, reaction jets for guidance, and rocket propulsion. These additional complexities will require substantial computational resources to model accurately and in a timely manner. It appears that only through highly efficient utilization of scalable computers will the capability to adequately address the full scope of these problems be realized.

Acknowledgment

Sponsored by ARO, ARPA, NASA-JSC, and by the Army High Performance Computing Research Center under the auspices of the Department of the Army, Army Research Laboratory cooperative agreement number DAAH04-95-2-0003/contract number DAAH04-95-C-0008. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred. CRAY C90 time and support for the third author were provided in part by the Minnesota Supercomputer Institute.

References

1. Aliabadi, S.K. and Tezduyar, T.E., "Parallel fluid dynamics computations in aerospace applications", *International Journal for Numerical Methods in Fluids*, **21** (1995) 783-805.
2. Tezduyar, T., Aliabadi, S., Behr, M., Johnson, A., Kalro, V. and Waters, C., "3D simulation of flow problems with parallel finite element computations on the CRAY T3D", in *Computational Mechanics'95, Proceeding of International Conference on Computational Engineering Science*, Mauna Lani, Hawaii (1995).
3. Tezduyar, T., Aliabadi, S., Behr, M., Johnson, A. and Mittal, S., "Massively parallel finite element simulation of compressible and incompressible flows", *Computer Methods in Applied Mechanics and Engineering*, **119** (1994) 157-177.
4. Sturek, W.B., "An experimental investigation of the supersonic turbulent boundary layer in a moderate adverse pressure gradient. Part I. A detailed description of the experimental and data tabulation", BRL Report No. 1506, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, 1970.

BUILDING SIMULATIONS FOR VIRTUAL ENVIRONMENTS AND PROTOTYPING

Ming C. Lin*
U.S. Army Research Office
Mathematical and Computer Sciences Division
4300 S. Miami Blvd., P.O. Box 12211
Research Triangle Park, NC 27709-2211

Abstract

In this paper, we describe a dynamic simulation system which provides accurate collision detection and response at interactive rates. Our collision detection algorithms utilize the properties of “locality” and “coherence”. As for collision response, we use a simple yet efficient and accurate “impulsed-based scheme” to compute the dynamics necessary to generate an appropriate response according to Newtonian physics.

1 Introduction

Over the last few years there has been a great deal of interest in the generation of real-time, 3D interactive virtual environments. This is largely due to training, acquisition and testing, entertainment, simulation-based design and other applications afforded by this new technology.

A major component of all these virtual environment systems is a real-time visual simulation system. The simulation takes in the geometry of the objects in the environment, their moving parts and physical properties to mimic their physical behaviors, when interacting with the user and other objects in the synthetic environment. Virtual objects, including the user’s image, need to behave like the real ones as in the physical world, in order to enhance the degree of realism. However, real time dynamic simulation on desk-top workstations has been unattainable till recently, due to the absence of efficient, practical collision detection algorithms and fast, accurate dynamics computations.

Other than virtual environments, a real-time dynamic simulation system is also an integral part of robotics and automation, molecular modeling and drug design, engineering analysis, and virtual prototyping. We will describe a dynamics simulation system which provides accurate collision detection and response at interactive rates.

The rest of the paper is organized as follows. In Section 2, we review some of the previous work in collision detection and response. Section 3 gives an overview on our simulation system and the underlying algorithms. We describe our novel approaches to the problem of collision detection and impact response in Section 4 and Section 5 respectively. Next, we demonstrate our experimental results on walkthrough environments and other applications in Section 6.

2 Previous Work

There is a rich body of literature in both analyzing the geometric contacts of moving objects and the motion dynamics of rigid bodies. Collision detection has been a fundamental problem in robotics, computational geometry, animation and simulation, physical-based modeling, molecular modeling. In virtual environments or physically-based modeling, where the motion is subject to dynamic constraints or external forces and cannot typically be

*Also with Computer Science Department, University of North Carolina, Chapel Hill, NC 27599-3175; lin@cs.unc.edu

expressed as a closed form function of time, various algorithms have been proposed [1,2,4,5,7,12]. At the same time, the emphasis in the computational geometry has been on the theoretically efficient intersection detection algorithms [10]. Most of them are restricted to a class of static objects and non-trivial to implement. Different methods have also been proposed to overcome the bottleneck of $O(N^2)$ pairwise tests for N moving objects in an environment. The simplest of these are based on spatial subdivision. Another approach operates directly on four-dimensional volumes swept out by object motion over time [6]. None of these approaches adequately address the issue of real-time, exact collision detection for virtual environments, which requires performance at interactive rates for thousands of moving objects.

Routh's seminal work is highly regarded and often considered as the foundation of rigid body dynamics [11]. Some of the earliest treatments of dynamics simulations can be found in [5,7]. Hahn also pioneered the "impulse-based" approach to model contact forces. However, many of these approaches suffered from inefficient collision detection algorithms. Newton [3] is often considered the most commonly used, general-purpose dynamic simulator in the academic research community. Its contact modeling is fairly simple and uses constrained-based dynamics for linkages. Recently, Baraff, Witken and others have also addressed this problem [1,2,13,14]

3 System Overview

Collision detection and dynamic simulation has a long history. It has been considered in robotics, animation, computational geometry, computer graphics, and now virtual reality literatures. One of major bottlenecks in building real-time dynamic simulators is finding a practical, efficient and simple collision detection algorithm [5]. Most common methods use embedded bounding boxes to check for interference or exhaustive checking. There are also algorithms in computational geometry which solves the intersection problems in linear time. However, most of these algorithms are nearly impossible to implement and subject to numerical robustness issues and geometric degeneracies. The other problem is to model the real-world physics and numerically simulate the contact and external forces at interactive rate.

We describe the system architecture of our dynamic simulator which addresses some of the critical issues in achieving real-time simulations. Our simulator consists of two main components: collision detection and impact response. Our collision detection algorithms utilize the properties of "locality" and "coherence". Several concepts from computational geometry and geometric modeling are incorporated to solve the collision detection problems among all geometric models in real time, for most of cases. The algorithms are generally applicable, but especially well suited to dynamic domains where objects are moving at small, discrete steps.

The performance of our collision detection for convex polyhedral models is independent of the model complexity and is a function of the object motion. The run time for checking interference between non-convex objects and curved models (such as NURBS, algebraic surfaces) is output sensitive. We also use a *sweep and prune* technique to reduce the quadratic number of pairwise interference tests in large simulation environments.

As far as collision response is concerned, at the current stage we use a simple yet efficient and accurate *impulsed-based scheme* to compute the dynamics necessary to generate an appropriate response according to Newtonian physics. Contact analysis is performed and collision response is calculated to display the motion change of colliding objects upon impacts. The resulting information is then sent into graphics pipeline for interactive display and rendering.

We also maintain a priority queue (implemented as a heap) of all pairs of objects that might collide. They are *adaptively* sorted by lower bound on time to collision; with the one most likely to collide (i.e. the one that has the smallest approximate time to collision) appearing at the top of the heap. Since the approximation is a *conservative* lower bound on the time to collision, no collisions are missed. Non-convex objects, which are represented as hierarchy trees are treated as single objects from the point of view of the queue.

4 Collision Detection

In most computer simulation systems, interactions among objects are generated by modeling the contact constraints and impact dynamics. Since prompt recognition of possible impacts is a key to successful response to collisions in a timely fashion, a simple and efficient algorithm for collision detection is necessary to fast and realistic simulation of moving objects.

The heart of our collision detection algorithm is a simple and fast incremental method [8] to compute the distance between two convex polyhedra. It utilizes convexity to establish “local applicability criteria” using Voronoi regions for verifying the closest feature pairs. As the objects travel through space, the algorithm takes advantage of coherence to keep track of the closest feature pairs between them to update their spatial relation. A preprocessing procedure is used to guarantee expected constant time performance when checking for contacts in a dynamic environment.

As for non-convex objects, we assume that each nonconvex object is given as a union of convex polyhedra or is composed of several nonconvex subparts, each of these can be further represented as a union of convex polyhedra or a union of concave subparts. We use a sub-part hierarchy tree to represent each nonconvex object. At each time step, we examine the possible interference by a recursive algorithm. The algorithm will only signal a collision if there is actually an impact between the sub-parts of two objects; otherwise, there is no collision between the two objects. This approach guarantees that we find the earliest collision between concave objects while reducing computation costs.

Furthermore, in order to avoid unnecessary computations and to speed up the run time, we use a scheduling scheme to reduce the frequency of collision detection. The algorithm maintains a priority queue (implemented as a heap) of all pairs of objects that might collide. They are sorted by “wake-up time” which is based on approximate time to collision; with the one most likely to collide appearing at the top of the heap. The approximation is a lower bound on the time to collision and is calculated adaptively, so no collisions are missed.

In addition to the priority scheduling scheme, we also use a *sweep and prune* approach to cull out objects which are far apart, based on their geometry and position. This method is simple and efficient. It is output sensitive and its run time is linearly dependent on the number of objects in the environment instead of quadratic dependence. We use dynamic bounding boxes, linear sweep and sort, and geometric coherence to quickly reject the object pairs, that are unlikely to collide within the next frame. This mechanism has reduced the number of pairwise tests dramatically, especially in a large walkthrough environment.

5 Collision Response

Given the ability to detect collisions at interactive rate, the applications also involve generation of a suitable collision response. The central problem in collision response is the computation of collision impulses. Accurate computation of impulses arising between colliding, sliding, rolling bodies is critical to the physical accuracy of the impact response.

There are two major approaches in computing collision impulses: the constrained-based method and the impulse-based scheme. Each has its own advantages. Constrained-based approach [1,2,14] is suited to model the motion constraints imposed by other contacting bodies. A perfect example is modeling a hinge joint or linkage bodies (such as human motion dynamics). On the other hand, impulse-based method is simple and easy to implement. In addition, impulse-based approach does not need to keep track of contact states (resting, rolling, sliding, colliding, etc.) and it presents a unified computational model for computing the collision impulses for impact response.

Impulse-based dynamics [5,9] assumes no explicit constraints on configuration of the moving objects. When the objects are not in contact, they are traveling in ballistic trajectory. Otherwise, all modes of contacts are modeled via series of impulses applied to the objects, whether they are bouncing around, sliding through or resting on top of each other. Obviously, this approach is very “collision intensive”, i.e. the simulator will need to check for possible contacts frequently. However, since we have extremely efficient algorithms for collision detection, we can compute the motion dynamics of rigid bodies in nearly real time on current graphics workstation.

6 Application Demonstration

We have tested the simulator on numerous cases including: interaction with architecture walkthrough, dynamics of threaded screw insertion, interlocked toroidal chain motion, bowling, non-holonomic motion of a ball on a spinning platter, etc. One of our goals is to eventually integrate our dynamic simulator in a fighting vehicle walkthrough system for training purpose. In such a large-scale environment, we have thousands, even millions of complex objects and we need to simulation interaction among the user and the virtual world. The complexity of *mega* models poses a grant challenge to real-time simulation. We have been collaborating with the Walkthrough Group at UNC Chapel Hill to integrate this system into their architecture walk-through and a multi-body simulation environment.

Here are some snapshots from our simulations. The preliminary results have shown great promises in achieving interactive dynamic simulation on the current graphics workstations. For more information and demonstration of our algorithms and systems, please refer to

<http://www.cs.unc.edu/~geom>

for a catalog of videos, examples, and technical reports.

Acknowledgements

We would like to thank Dinesh Manocha, Jonathan Cohen, Krish Ponamagi and the Walkthrough Group at the University of North Carolina, Chapel Hill for their collaboration and the architecture model for demonstration. We are also grateful to Brian Mirtich for his help in implementing parts of the system and John Canny for his constructive feedback.

References

- [1] Baraff, D., *Analytical Methods for Dynamic Simulation of Non-Penetrating Rigid Bodies*, ACM Computer Graphics, 23 (3): pp. 223-232, July 1989.
- [2] Baraff, D., *Curved Surfaces and Coherence for Non-Penetrating Rigid Body Simulation*, ACM Computer Graphics, 24 (4): pp. 19-28, 1990.
- [3] Cremer, J. F. and Stewart, A. J., *The Architecture of Newton, A General Purposed Dynamic Simulator*, Proceedings of the IEEE International Conference on Robotics and Automation, May 1994.
- [4] Garcia-Alonso, A., Serrano, N., and Flaquer, J., *Solving the Collision Detection Problem*, IEEE Computer Graphics and Applications, 13 (3): pp. 36-43, 1994.
- [5] Hahn, J.K., *Realistic Animation of Rigid Bodies*, ACM Computer Graphics, 22 (4): pp. 299-308, 1988.
- [6] Hubbard, P.M., *Interactive Collision Detection*, Proceedings of IEEE Symposium on Research Frontier in Virtual Reality, October 1993.
- [7] Moore, M. and Wilhelms J., *Collision Detection and Response for Computer Animation*, ACM Computer Graphics, 22 (4): pp. 289-298, 1988.
- [8] Lin, Ming C., *Efficient Collision Detection for Animation and Robotics*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, University of California, Berkeley, 1993.
- [9] Mirtich, B. and Canny, J., *Impulse-Based Simulation of Rigid Bodies*, ACM Symposium on Interactive 3D Graphics, pp.181-188, 1995.
- [10] Preparata, F.P. and Shamos, M.I., *Computational Geometry*, Springer-Verlag, New York, 1985.
- [11] Routh, E. J., *Elementary Rigid Dynamics*, 1905.
- [12] Thibault, W. and Naylor, B., *Set Operations on Polyhedra Using Binary Space Partitioning Trees*, ACM Computer Graphics, 4, 1987.
- [13] Wang, Y. and Mason, M., *Modeling Impact Dynamics for Robot Operations*, Proceedings of the IEEE International Conference on Robotics and Automation, pp. 678-685, May 1987.
- [14] Witkin, A., Gleicher, M., and Welch, W., *Interactive Dynamics*, ACM Computer Graphics, 24 (2): pp. 11-22, March 1990.

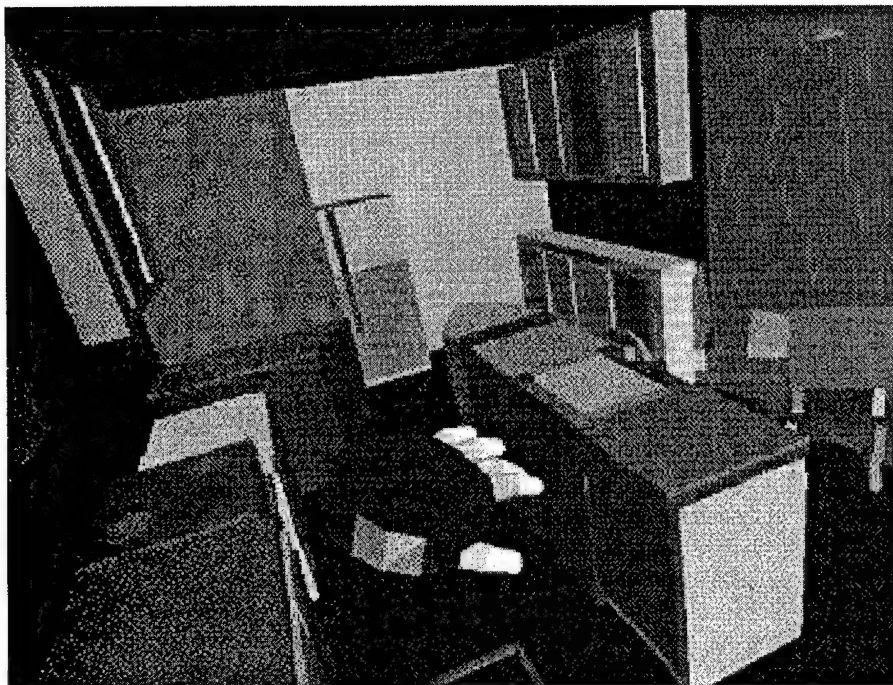


Figure 1: A user hand image moves through a kitchen walkthrough environment

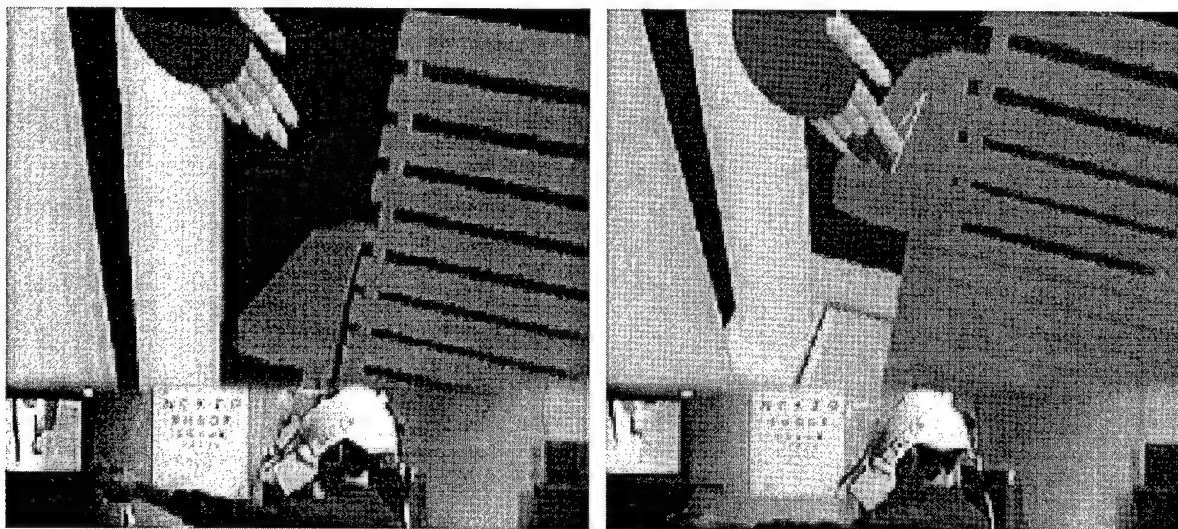


Figure 2: A split screen scenario shows both the virtual world (top), where the hand touches a virtual swing in the porch, and the actual scene (bottom), where the user is immersed in the synthetic environment.

Chemical and Biological Sciences

THEORETICAL DETERMINATION OF POTENTIAL HAZARDS IN THE HANDLING OF CK AND AC MUNITIONS

Betsy M. Rice*, Sharmila V. Pai and Cary F. Chabalowski

U. S. Army Research Laboratory

Aberdeen Proving Ground, Maryland 21005-5066

Abstract

Quantum mechanical calculations [*ab initio* and Density Functional Theory (DFT)] were performed to investigate reaction mechanisms for formation and decomposition of the six-membered rings $C_3N_3X_3$, where X denotes Cl or H. $C_3N_3Cl_3$ (cyanuric chloride) and $C_3N_3H_3$ (*sym*-triazine) are the polymerization products of agents CK (CICN) and AC (HCN). The process of the polymerization to the trimer species of CK and AC has long been rumored to be highly exothermic and thought to be the cause of explosions of cylinders containing these agents. The calculations presented here determine energy release upon reaction, activation barriers and details of the reaction mechanisms for the polymerization of these agents. The calculations investigate concerted triple association and step-wise addition reactions leading to the trimers, and provide information on the low-energy pathway to polymerization. *Ab initio* MP2 and non-local DFT geometry optimizations for critical points on the CK and AC potential energy surfaces were calculated with the 6-311+G* and cc-pVTZ basis sets, respectively. The MP2 calculations were followed by QCISD(T) energy refinements for the AC system. Good agreement with experiment is found for the predicted geometries and frequencies of *sym*-triazine, HCN, CICN and cyanuric chloride. All calculations indicate that CK and AC have similar features on the potential energy surfaces for the trimerization reactions. All calculations show that the lowest-energy formation mechanism for trimerization is the concerted triple association. In both chemical systems, a weakly-bound $(XCN)_3$ cluster is a pre-reaction intermediate to the trimerization. This reduces entropic hindrance to the unusual concerted triple association of the CK and AC molecules. Both systems have a large energy release upon trimerization and similar barriers to activation of the reactions. The energy available to the surrounding system upon crossing the reaction barrier is sufficient to initiate chain reactions. Preliminary studies on polymerization reactions of AC and CK to the tetramers predict significant energy release that could initiate additional reactions.

1. Introduction

Agents CK (CICN) and AC (HCN) were standard and substitute fill, respectively, of chemical munitions during World War II. Excavation of abandoned chemical warfare munitions dumps for environmental cleanup have uncovered old munitions, some of which are assumed to contain CK or AC. Due to the massive production of these munitions in the 1940's, it is expected that more CK and AC munitions will be uncovered in old disposal sites. These cyanide compounds are extremely toxic and must be destroyed. However, there are reports that CK and AC can undergo violent reactions, and that cylinders containing these agents have been reported to explode.¹ It is crucial that the chemistry behind the initiation of these violent events be characterized before handling these munitions as part of an environmental cleanup.

It has long been known that CK will spontaneously polymerize upon long periods of containment.¹ One of the primary polymerization products of CK is the trimer, cyanuric chloride. However, no studies have been undertaken to determine heats of reaction and activation barriers for this polymerization reaction. AC, on the other hand, is well-known to release substantial energy (explosive) upon polymerization, presumably to its trimer, *sym*-triazine.¹ Although the chemical simplicity of AC and its multimers has made them attractive candidates for experimental and theoretical study, details of the polymerization reactions of AC have not been determined. To explore the possible explosivity of these cyanide compounds, more information is needed on the reaction mechanisms and barrier heights to formation from XCN , $X=Cl$ or H .

During the course of this work, samples of the solid contents of an unidentified bomb located at the Dugway Proving Ground were shown to consist of over 80% polymer of CK. 42% was identified as cyanuric chloride and 41% was found to have the same mass weight as a tetramer of CK. Since no reactive activity was observed, indicating that the crystal was not sensitive, we have focused our studies on examining the process of polymerization of the monomer species as the energy-release step that could initiate the reported explosions.

The only experimental observation pertinent to these reactions is the study of the decomposition of *sym*-triazine to form HCN² (the reverse of the polymerization reaction of AC)



Photodissociation experiments at 248 and 193 nm showed that *sym*-triazine decomposes in a concerted manner only and forms three HCN molecules.² These measurements provided an upper bound of the barrier to reaction (1) (115 kcal/mol). The results also indicate that the transition state for (1) has three-fold symmetry.² Evidence of a step-wise decomposition reaction



was not observed.²

Concerted triple association of three XCN molecules to form *sym*-triazine or cyanuric chloride seems more improbable than the concerted triple dissociation reaction in (1). In such an association, the entropic effects associated with bringing three XCN molecules together in a concerted fashion, especially in the gas phase, would seem so large as to prohibit reaction. However, a pre-reaction intermediate with a structure favorable to concerted triple association would reduce this entropic hindrance. Experiment has shown the existence of a cyclic hydrogen-bonded cluster (HCN)₃.³ It is conceivable that contained XCN gas could form this weakly bound cyclic trimer which would be in a sterically favorable arrangement for a concerted association:



Reaction energy appropriately imparted to the cluster would allow Reaction (3) to occur. In this work, we report results of our investigations of Reactions (1)-(3).

There are numerous isomers of the tetrameric forms of AC and CK; however, the results presented here will focus on the concerted quadruple association reaction to form the cyclic X₄C₄N₄ species via formation of a weakly-bound cyclic (XCN)₄ cluster:



We are in the process of studying other reaction mechanisms for formation of tetramers of CK and AC.

2. Methods

All calculations reported in this work were done using the Gaussian 94 set of programs.⁴ Critical points on the CK and AC potential energy surfaces corresponding to reactions (1)-(3) were determined through MP2 and DFT geometry optimizations and characterized through normal mode analyses. The MP2 calculations used the 6-311+G* and cc-pVTZ basis sets⁴ for the CK and AC systems, respectively. The DFT calculations used the cc-pVTZ basis set⁴ for both systems. QCISD(T) energy refinements were calculated for each of the MP2-optimized structure in the AC system. Although several density functionals were used to evaluate these potential energy surfaces, the B3LYP functional⁴ (Becke's three parameter hybrid exchange functional with the LYP correlation energy functional) provided best agreement with experiment² and highest level *ab initio* calculations.^{5,6} Therefore, we report only the B3LYP results. All calculations used default options and met default convergence criteria of Gaussian 94.⁴ Intrinsic Reaction Coordinate (IRC) calculations were performed at the MP2 and DFT/6-31G** level to establish reaction paths. Saddle points for Reaction (2) were not determined for AC (discussed below); however, the transition state for formation of the dimer in Reaction (2) was determined for CK. B3LYP/6-31G** calculations were performed for points associated with Reaction (4).

3. Results and Discussion

In this section, we report overall reaction energetics for the polymerization of AC and CK to their corresponding six- and eight-membered rings. Details on the molecular properties of AC can be found in Refs. 5 and 6.

3.1. 3HCN \rightarrow *sym*-triazine

Molecular structures for *sym*-triazine, the dimer $\text{H}_2\text{C}_2\text{N}_2$, the $(\text{HCN})_3$ cluster, the HCN molecule, and the transition state for Reactions (1) and (3) are shown in the energy schematic below (fig. 1). The figure lists the zero-point corrected energies of critical points on the AC potential energy surface relative to the reactant, 3 HCN molecules. Our calculations predict substantially smaller barriers than the experimentally determined upper bound (72 kcal/mol).² The B3LYP prediction of the heat of reaction is in closer agreement with experiment (-43.2 kcal/mol) than the *ab initio* value.⁵ The reactant minimum for both mechanisms, corresponding to three isolated HCN molecules, is located at the center of the figure. The reaction that proceeds to the right from the 3 HCN minimum corresponds to the concerted triple association reaction. The reaction that proceeds toward the left from the 3 HCN minimum corresponds to the step-wise addition reaction. This figure shows that the barrier to concerted triple association is lower than the step-wise addition reaction by at least 6 kcal/mol. The existence of the $(\text{HCN})_3$ cluster on the reaction path for the triple concerted association reaction reduces the entropic hindrance to this unusual reaction mechanism. The barriers to formation of and insertion of HCN into the dimer species were not calculated, because the minimum corresponding to dimer + HCN is higher in energy than the barrier to the concerted dissociation. Even if there were no barriers to formation of or insertion into the dimer, the formation of this complex requires more energy than needed to cross the barrier to the triple association. The results also show that once the barrier to formation is crossed, over 80 kcal/mol energy is available to the surrounding system. This available energy is sufficient to initiate a second reaction, possibly leading to explosion.

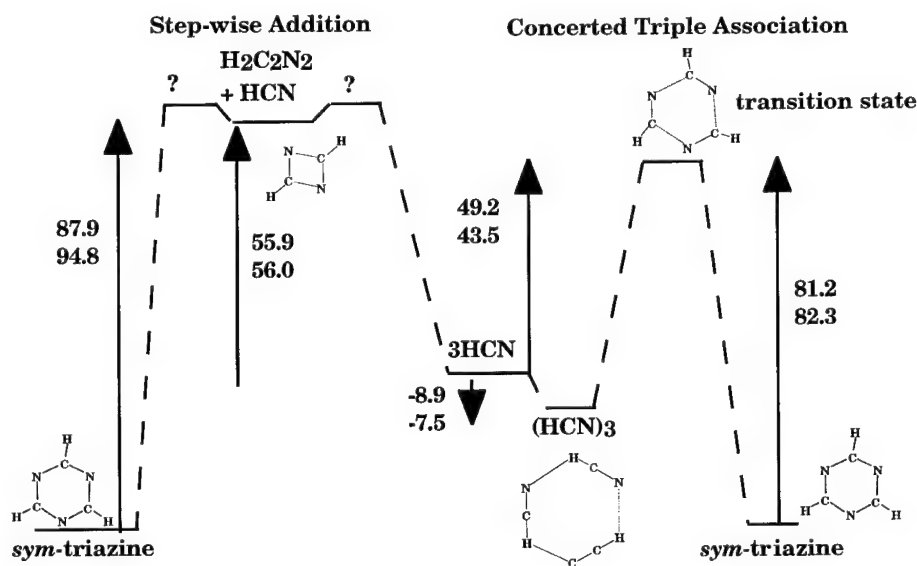


Figure 1. Zero-point corrected relative energies (kcal/mol) of critical points on the *sym*-triazine (AC) potential energy surface. Our best *ab initio* [QCISD(T)//MP2/cc-pVTZ] and DFT (B3LYP/cc-pVTZ) predictions are shown. The DFT value is listed below the *ab initio* prediction.

3.2. 3CICN--> cyanuric chloride

Molecular structures for cyanuric chloride, the dimer $\text{Cl}_2\text{C}_2\text{N}_2$, the $(\text{CICN})_3$ cluster, the CICN molecule, and the transition state for Reactions (1) and (3) are shown in figure 2. The structural parameters are very similar to the corresponding hydrogen analog from the AC system, with the exception of the $(\text{CICN})_3$ cluster. Its structure is somewhat skewed as compared to the more spherical $(\text{HCN})_3$ cluster. Subsequent normal mode analyses characterized each species. Figure 2 also lists the zero-point corrected energies of critical points on the CK potential energy surface relative to the reactant, 3 CICN molecules. The *ab initio* prediction of the barrier to concerted triple association is significantly smaller than for the AC system; it predicts a barrier of 38.4 kcal/mol whereas the corresponding AC value is 49.2 kcal/mol. However, the B3LYP prediction of the barrier, 42.9 kcal/mol, is remarkably similar to that of the AC system (43.5 kcal/mol). The B3LYP predictions for this system have the advantage of being calculated with the largest basis set (cc-pVTZ). As for the AC system, the existence of the $(\text{CICN})_3$ cluster on the reaction path for the triple concerted association reaction reduces the entropic hindrance to this unusual reaction mechanism.

A significant difference between the AC and CK systems is the energy of the dimer + CICN species relative to three isolated CICN molecules. The minimum corresponding to the dimer is lower in energy than the barrier to the concerted triple association. Therefore, we calculated the barriers to formation of the dimer species. We found that the barrier to formation of the dimer + CICN is higher in energy than the barrier to concerted triple association by over 20 kcal/mol. Thus, the step-wise addition reaction is eliminated as the low-energy pathway making it unnecessary to calculate the barrier to insertion of a third CICN molecule into the dimer.

The energy release upon crossing the barrier to triple association is larger for the CK system than for the AC system by 15 to 20 kcal/mol. Therefore, as for the AC system, once the saddle point is crossed, sufficient energy to initiate a second reaction is released, and could explain initiation of chain reactions leading to explosion for this system.

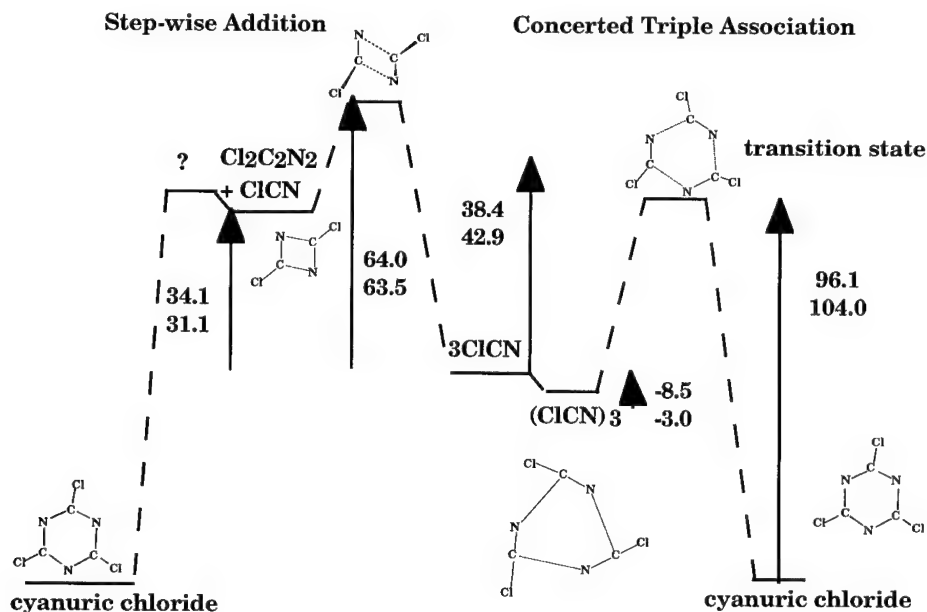


Figure 2. Zero-point corrected relative energies (kcal/mol) of critical points on the CK potential energy surface. Our best *ab initio* (MP2/6-311+G*) and DFT (B3LYP/cc-pVTZ) predictions are shown. The DFT value is listed below the *ab initio* prediction.

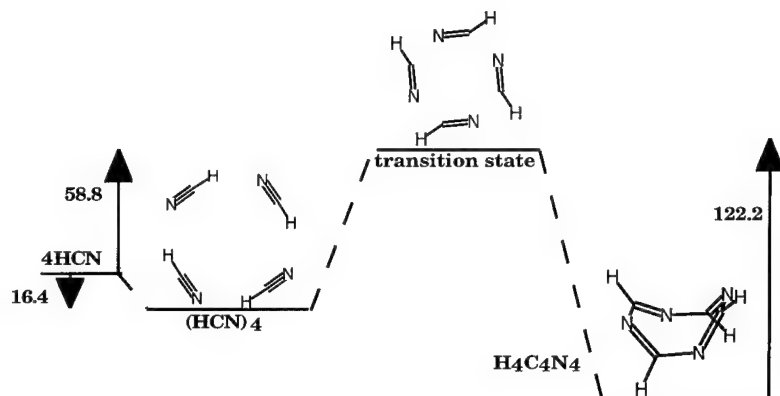
3.3. $4\text{HCN} \rightarrow \text{X}_4\text{C}_4\text{N}_4$ 

Figure 3. B3LYP/6-31G** zero-point corrected energies (kcal/mol) of critical points along the AC concerted quadruple association reaction relative to isolated HCN.

We have calculated the barriers to and energy release upon formation of an AC tetramer from isolated HCN at the B3LYP/6-31G** level. Figure 4 shows the existence of a hydrogen-bonded cyclic $(\text{HCN})_4$ cluster and a substantial barrier to formation of the covalently bound boat-shaped tetramer species. Once the barrier is crossed, 122.2 kcal/mol is released to the surroundings. We have also determined the heat of reaction of CK tetramer (-70.3 kcal/mol); however, we have not located the barrier to formation at this time. If we assume that the CK and AC tetramer systems behave in a similar manner as the trimer systems, then the CK tetramer formation barrier will be of similar magnitude as the AC counterpart (~60 kcal/mol). Using this approximation along with the heat of reaction, we estimate that the energy available to the system upon crossing the barrier is ~125-135 kcal/mol. It is possible that step-wise addition reactions in the formation of the tetramer species might be lower than the concerted association reactions; we are investigating these reactions at this time.

Acknowledgements

This work was supported by the Program Manager Non-Stockpile Chemical Materiel U. S. Army Chemical Demilitarization and Remediation Activity. S. Pai is a National Research Council Fellow.

References

1. Herbert S. Aaron, "Potential Hazards in the Handling of Aged AC and CK Munitions: A Literature Review", ERDEC Technical Report, in press.
2. Ondrey, G. S. and Bersohn, R., "Photodissociation Dynamics of 1,3,5-triazine", *J. Chem. Phys.*, Vol. 81, No. 10, pp. 4517-4520, November 1984.
3. Jucks, K. W. and Miller, R. E., "Near Infrared Spectroscopic Observation of the Linear and Cyclic Isomers of the Hydrogen Cyanide Trimer", *J. Chem. Phys.*, Vol. 88, No. 4, pp. 2196-2204, February, 1988.
4. Gaussian 94, Revision B.1, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1995 and references therein.
5. Pai, S., Chabalowski, C. and Rice, B., "A Comparative Study of Nonlocal Density Functional Theory and *Ab Initio* Methods: The Potential Energy Surface of *sym*-triazine Reactions", *J. Phys. Chem.*, submitted.
6. Pai, S. V., Chabalowski, C. F. and Rice, B. M., "Ab Initio Study of Reactions of *sym*-Triazine", *J. Phys. Chem.*, Vol. 100, No. 14, xxx, April, 1996.

Carboxylesterase: Regulatory Control and Peptide-Induced Secretion of an Endogenous Scavenger for Organophosphorus Agents

Donald M. Maxwell*

U. S. Army Medical Research Institute of Chemical Defense
Aberdeen Proving Ground, MD, 21010-5425

Kenneth D. Lanclos
Medical College of Georgia
Augusta, GA, 30912-2100

Hendrik P. Benschop
TNO Prins Maurits Laboratory
Rijswijk, The Netherlands

Abstract

Current medical treatment against organophosphorus chemical warfare agents (i.e., nerve agents) is effective in protecting against lethality, but survivors of exposure to multiple lethal doses of nerve agents exhibit postexposure incapacitation. Our toxicokinetic studies of the nerve agent soman revealed that postexposure incapacitation was most likely due to the persistence of toxic levels of soman in blood for as long as 5 hr after exposure. Our approach to alleviating this incapacitation problem was to attempt to sequester soman with carboxylesterase (CaE), an endogenous scavenger for nerve agents. By carboxyl terminal mutagenesis of human CaE expressed in the HepG2 cell line and by comparison of carboxyl terminal sequences of CaE from different species, the retention signal that controls secretion of CaE from liver into plasma was identified as His-X-Glu-Leu. To competitively inhibit the retention signal receptor for CaE and stimulate secretion of CaE, a peptide with the sequence His-Ile-Glu-Leu was synthesized and evaluated for its ability to cause *in vitro* secretion of CaE from liver microsomes. Complete secretion of microsomal-bound CaE into the surrounding medium was achieved with 5 μ M peptide in 30 min at 25° C (pH 7.4). These observations suggest that a peptide drug should be able to effectively increase plasma levels of an endogenous scavenger for nerve agents to provide improved protection against these chemical warfare agents.

1. Introduction

Nerve agents are the most toxic chemical warfare agents that threaten U.S. military personnel. Estimated lethal dosages of nerve agents in a 70-kg human are < 1 mg. The acute toxicity of these organophosphorus (OP) agents is due to their irreversible inhibition of acetylcholinesterase (AChE), an enzyme that terminates the action of the neurotransmitter acetylcholine. By increasing acetylcholine levels in cholinergic synapses, inhibition of AChE produces a variety of adverse physiological and behavioral effects culminating in death, usually by respiratory failure. Current medical treatment consists of a regimen of anticholinergic drugs, such as atropine, to counteract the accumulation of acetylcholine and oximes, such as HI-6, to reactivate OP-inhibited AChE (Dunn and Sidell, 1989). The inability of oximes to provide adequate protection against some nerve agents stimulated the development of pretreatment with carbamates, such as pyridostigmine, in which carbamylation of AChE protects it against inhibition by nerve agents. Spontaneous decarbamylation of AChE after the nerve agent has been detoxified then generates enough active AChE to allow normal cholinergic neurotransmission. Although both oxime and carbamate antidotal regimens are effective in preventing lethality from nerve agents in nonhuman primates, survivors of exposure to multiple lethal doses of nerve agents exhibit postexposure incapacitation (Castro *et al.*, 1991). The purpose of our research was to explore the nature of OP-induced postexposure incapacitation and the possible use of an endogenous scavenger to alleviate this problem.

2. Toxicokinetics of Nerve Agents

To examine the toxicological state of OP-incapacitated animals, toxicokinetic studies were performed in atropinized, anesthetized, and artificially ventilated animals receiving 0.8 - 6 x the median lethal dose (LD_{50}) of soman, the nerve agent that is the most resistant to medical treatment. Soman (pinacolyl methylphosphonofluoridate) is very reactive in biological fluids and the development of specific sample stabilization procedures was required before soman could be quantitatively extracted and analyzed. This stabilization problem was solved by acidification of biological samples to stop enzymatic degradation and addition of aluminum sulfate to complex fluoride ions that would otherwise regenerate free soman from soman-inhibited esterases. An additional problem was that soman consists of four stereoisomers that vary widely in their ability to inhibit AChE and in their toxicity (Benschop *et al.*, 1984). An exceptional level of analytical sensitivity was needed to measure *in vivo* concentrations of the most toxic stereoisomers, such as C(-) P(-) and C(+) P(-) soman, because their LD_{50} values were $< 5 \mu\text{g/kg}$.

Analytical techniques were developed that not only had sufficient sensitivity to measure *in vivo* concentrations of soman but also had the ability to differentiate between the various stereoisomers. Ethyl acetate extracts of stabilized blood samples were analyzed by gas chromatography with an alkali flame ionization detector after thermodesorption/cold trap injection. Stereoisomers of soman could be resolved on an optically active Chirasil-Val column with a minimum detectable concentration of 1.5 pg/ml (8.3 pM) for each stereoisomer (Benschop *et al.*, 1991).

Toxicokinetic analysis of blood concentrations of soman revealed that the terminal elimination rate of soman was much slower in animals receiving 6 x LD_{50} of soman than in animals receiving 1 x LD_{50} (fig. 1). Therefore, concentrations of 150 pM soman, the lowest concentration that is capable of inhibiting AChE, were maintained in animals that received 6 x LD_{50} of soman for as long as 317 min in rats, 126 min in guinea pigs and 74 min in monkeys (table 1). Inasmuch as the biological half-lives for oximes are < 30 min and for carbamates are < 50 min, a major cause of postexposure incapacitation is most likely the lingering presence of soman after antidotes have been eliminated. This prolonged exposure to toxic concentrations of soman may also produce irreversible impairment, since neither continued administration of antidotes to maintain therapeutic concentrations nor the administration of the anticonvulsant diazepam eliminates postexposure incapacitation in nonhuman primates (Castro *et al.*, 1991).

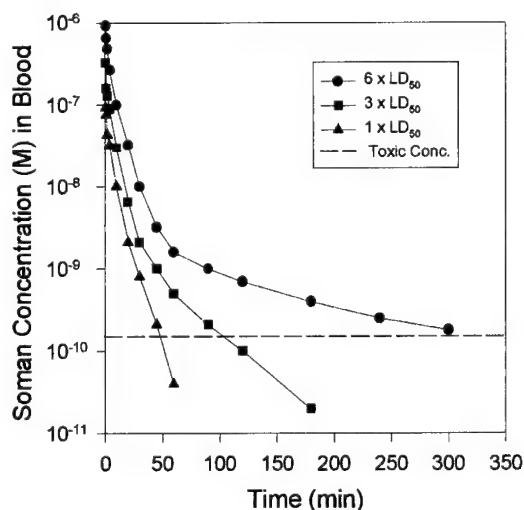


Figure 1. Toxicokinetics of soman. The time-course of C(\pm) P(-) stereoisomers of soman in blood was measured in atropinized, anesthetized, and artificially ventilated rats after 1 x, 3 x, or 6 x LD_{50} i.v. of soman. The toxic concentration of soman (--) was the lowest concentration (150 pM) that will inhibit AChE. Data taken from Benschop *et al.* (1991).

Table 1. Effect of Soman Dose on Duration of Toxic Concentration of Soman Stereoisomers in Blood

Species	Duration (min) of Toxic Level of Soman ^a		
	0.8 - 1 x LD_{50} ^b	2 - 3 x LD_{50}	6 x LD_{50}
Rat ^c	37	95	317
Guinea Pig ^d	11	104	126
Monkey ^c	--	49	74

a. Lowest concentration of C(\pm) P(-) stereoisomers of soman (150 pM) that will inhibit AChE.

b. Multiples of LD_{50} administered i.v. to animals. Values for 1 x LD_{50} of C(\pm) P(-) stereoisomers of soman in rat, guinea pig, and monkey were 41, 14, and 5 $\mu\text{g/kg}$, respectively.

c. Data taken from Benschop *et al.* (1991).

d. Data taken from Due *et al.* (1993).

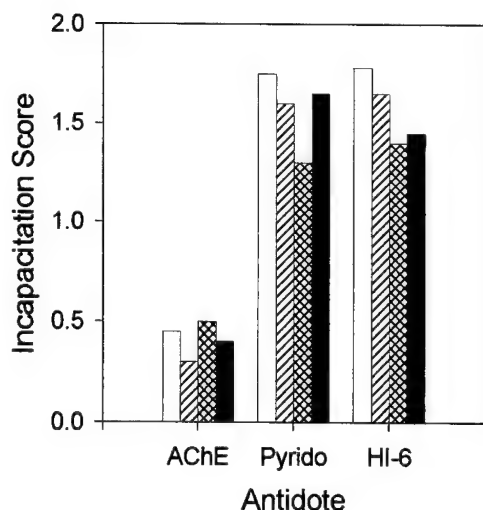


Figure 2. Effect of antidote regimen on soman-induced incapacitation in mice. Pyridostigmine with atropine, HI-6 with atropine, and AChE were tested against $6 \times \text{LD}_{50}$ of soman. Incapacitation was identified by increased lacrimation (\square), convulsions ($//$), reduced activity level (x), and behavioral deficits (\blacksquare). Data taken from Maxwell *et al.* (1993).

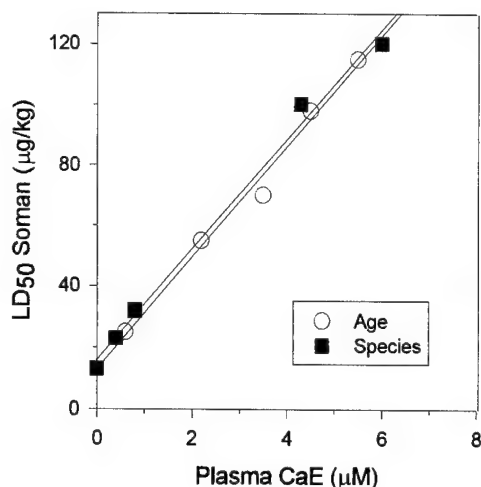


Figure 3. Effect of plasma CaE concentration on soman LD_{50} s.c. in different species and aging rats. CaE was measured by titration with ^{14}C -soman (Little *et al.*, 1990). Data points (from lower left to upper right of graph) for species were monkey, rabbit, guinea pig, rat and mouse and for ages of rats were 15, 120, 90, 30 and 60 days. Data taken from Maxwell *et al.* (1991).

3. Bioscavenger Protection Against Nerve Agents

Toxicokinetic studies have shown that bioscavengers can reduce the duration of *in vivo* toxic concentrations of soman (Due *et al.*, 1993). The therapeutic benefit of reducing postexposure soman levels below toxic concentrations was shown by the decrease in soman-induced incapacitation that we observed in scavenger-protected mice vs. oxime- or carbamate-protected mice (fig. 2). Although oxime protection with HI-6, carbamate protection with pyridostigmine, or scavenger protection with exogenous AChE produced equal protection against lethality from soman, scavenger protection produced 75% less postexposure incapacitation from soman as measured by lacrimation, convulsions, activity level, or behavioral deficits (Maxwell *et al.*, 1993).

A major problem with scavenger protection by administration of an exogenous enzyme, such as AChE, is that its large molecular weight requires the injection of several grams of protein to sequester lethal doses of OP agents in humans, which also increases the possibility of an adverse immune response to the scavenger. To avoid these problems our approach to scavenger protection against nerve agents has focused on the regulatory control of an endogenous scavenger, carboxylesterase (CaE), whose concentration in plasma is the major determinant of variation in soman toxicity (Maxwell *et al.*, 1991). For example, the 5-fold variation in soman LD_{50} values as rats aged and the 10-fold variation in soman LD_{50} values in different species were both directly proportional to the variation in their respective plasma concentrations of CaE (fig. 3).

4. Regulatory Control of CaE Secretion

Although the importance of plasma CaE in modulating soman toxicity with age or species is well established, the tissue source and regulatory control of this detoxification enzyme is poorly defined. We have examined the pattern of CaE levels in liver and plasma across six mammalian species by enzymatic analysis with specific substrates and inhibitors (Maxwell, 1992) as well as active site titration with radiolabelled soman (Little *et al.*, 1990). Species with high plasma levels of CaE (i.e., mouse and rat) had equal or lower levels of CaE in liver, and species with low plasma levels of CaE (i.e., human, hamster, rabbit and swine) had high CaE levels in liver, which suggested that plasma levels of CaE resulted from the secretion of CaE from liver into plasma (fig. 4).

Recent investigations of the regulatory mechanisms of protein secretion have concluded that there are intrinsic membrane

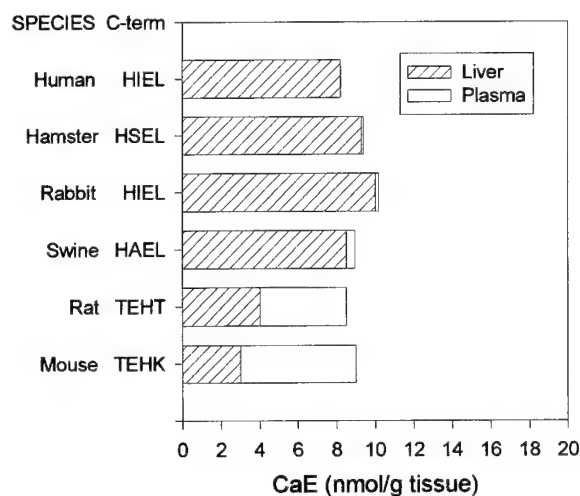


Figure 4. Concentration of CaE in liver and plasma of different species. CaE was measured by titration with ^{14}C -soman (Little *et al.*, 1990) and comparisons were based on 1 g of liver and 1 ml of plasma. The amino acid sequences of the carboxyl terminal (C-term) of liver CaE from different species were taken from Gentry and Doctor (1995).

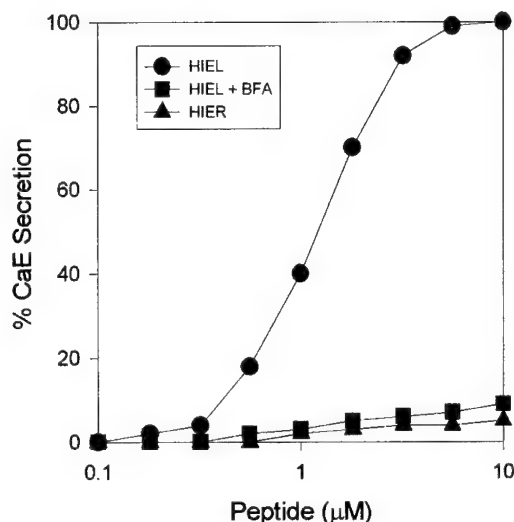


Figure 5. Peptide-induced secretion of CaE from rabbit liver microsomes. A suspension of liver microsomes was incubated with each peptide for 30 min at pH 7.4 (25° C) and centrifuged at 100,000 x g for 90 min to separate microsomal CaE from secreted CaE. The His-Ile-Glu-Leu peptide (HIEL) was tested in the presence and absence of 1 $\mu\text{g}/\text{ml}$ brefeldin A (BFA).

receptors for carboxyl terminal sequences of proteins that are located in the endoplasmic reticulum, where protein synthesis occurs. Proteins containing specific carboxyl terminal sequences are bound to these receptors while proteins containing other sequences are transported to the Golgi for secretion (Pelham, 1990). To examine whether such a mechanism may control the secretion of CaE, human liver CaE (hCaE) was cloned and expressed in a eukaryotic system. The cDNA for hCaE was cloned into plasmids containing the simian virus 40 (SV40) enhancer/promoter for transient expression in the HepG2 liver cell line. The plasmid-encoded product exhibited the same substrate and inhibitor specificity and isoelectric focusing profile as naturally occurring hCaE and was normally retained in the HepG2 cells. However, hCaE was secreted into the medium when its carboxyl terminal leucine was mutated to arginine, which established the importance of the carboxyl terminus in regulating the secretion of hCaE (data not shown). To determine the complete carboxyl terminal retention signal, we compared the carboxyl terminal sequences of liver CaE from four species (human, hamster, rabbit and swine) in which little CaE is secreted into the plasma with two species (rat and mouse) in which extensive CaE secretion into plasma occurs (fig. 4). Species in which CaE was generally retained in the liver contained the consensus carboxyl terminal sequence His-X-Glu-Leu (HXEL), where X was a variable amino acid. This carboxyl terminal sequence was not found in species in which CaE was generally secreted, which suggested that it constituted the complete CaE retention signal.

5. Peptide-Induced Secretion of CaE

To competitively inhibit the retention signal receptor for CaE and stimulate secretion of CaE, a peptide containing the sequence His-Ile-Glu-Leu (HIEL) was synthesized and evaluated for its ability to cause *in vitro* secretion of CaE from crude liver microsomes. Microsomes that contained rough and smooth endoplasmic reticulum as well as Golgi were prepared by differential centrifugation of liver homogenates in 0.25 M sucrose (pH 7.4 at 4° C). Rabbit liver microsomes were used for these secretion investigations because rabbit and human CaE share the same carboxyl terminal sequence (See fig. 4) and presumably the same retention signal receptor specificity. Complete secretion of microsomal-bound CaE into the surrounding medium was achieved with 5 μM HIEL peptide in 30 min at 25° C (fig. 5). The specificity of this effect was demonstrated by the absence of stimulation of CaE secretion by closely related peptides, such as HIER, where leucine was replaced by arginine. The ability of brefeldin A, an inhibitor of Golgi function, to block the HIEL-induced increase in soluble CaE activity indicated that HIEL was stimulating Golgi-mediated

secretion of CaE and not merely elevating soluble levels of CaE by general disruption of microsomal structure.

6. Conclusion

The goal of medical protection against nerve agents is to protect against the lethality and incapacitation produced by exposure to 5 x LD₅₀ of these agents. Current medical treatments using a combination of anticholinergics, oximes, carbamates, and anticonvulsants are effective in preventing lethality at these exposure levels, but survivors exhibit postexposure incapacitation. Our studies have laid the foundation for the development of a simpler and more effective approach to protection against nerve agents. We have established that postexposure incapacitation is more easily alleviated by intervention with scavengers in the toxicokinetics of nerve agents than by pharmacological antagonism of nerve agent toxicity. We have overcome the disadvantages of exogenous scavengers (i.e., large size, adverse immune response) by developing approaches that enhance the secretion of CaE, an endogenous scavenger. By carboxyl terminal mutagenesis of cloned human CaE and by analysis of the species variation in carboxyl terminal sequences of CaE, we have identified the retention signal that controls the secretion of this endogenous scavenger into plasma, where it is most effective against OP agents. We have also synthesized a low molecular weight peptide that mimics the CaE retention signal and stimulates the secretion of CaE by competitively inhibiting the retention signal receptor. Thus, our new understanding of the regulatory control of CaE secretion has provided a range of new opportunities, from gene therapy to peptide drugs, for the development of improved protection against nerve agents through the use of endogenous scavengers.

References

1. Dunn, M.A. and Sidell, F.R., "Progress in Medical Defense Against Nerve Agents," *J. Amer. Med. Assoc.*, Vol 262, pp. 649-652, 1989.
2. Castro, C.A., Larsen, T., Finger, A.V., Solana, R.P. and McMaster, S.B., "Behavioral Efficacy of Diazepam Against Nerve Agent Exposure in Rhesus Monkeys," *Pharmacol. Biochem. Behav.*, Vol 41, pp. 159-164, 1991.
3. Benschop, H.P., Konigs, C.A.G., Van Genderen, J. and De Jong, L.P.A., "Isolation, Anticholinesterase Properties and Toxicity in Mice of the Four Stereoisomers of the Nerve Agent Soman," *Toxicol. Appl. Pharmacol.* Vol 72, pp. 61-74, 1984.
4. Benschop, H.P. and De Jong, L.P.A., "Toxicokinetics of Soman: Species Variation and Stereospecificity in Elimination Pathways," *Neurosci. Biobehav. Rev.*, Vol 15, pp. 73-77, 1991.
5. Due, A.H., Trap, C.H., Van Der Wiel, H.J. and Benschop, H.P., "Effect of Pretreatment with CBDP on the Toxicokinetics of Soman Stereoisomers in Rats and Guinea Pigs," *Arch. Toxicol.*, Vol 67, pp. 706-711, 1993.
6. Maxwell, D.M., Brecht, K.M., Doctor, B.P. and Wolfe, A.D., "Comparison of Antidote Treatment Against Soman by Pyridostigmine, HI-6 and Acetylcholinesterase," *J. Pharmacol. Exp. Therap.*, Vol 264, pp. 1085-1089, 1993.
7. Maxwell, D.M., Wolfe, A.D., Ashani, Y. and Doctor, B.P., "Cholinesterase and Carboxylesterase as Scavengers for Organophosphorus Agents," In: *Cholinesterases: Structure, Function, Mechanism, Genetics and Cell Biology*, Massoulie, J., Bacou, F., Barnard, E., Doctor, B. and Quinn, D.M., eds., American Chemical Society, Washington, DC, pp. 206-209, 1991.
8. Maxwell, D.M., "The Specificity of Carboxylesterase Protection Against the Toxicity of Organophosphorus Compounds," *Toxicol. Appl. Pharmacol.*, Vol 114, pp. 306-312, 1992.
9. Little, J.S., Maxwell, D.M., Fox-Talbot, M.K., Brecht, K.M. and Lenz, D.E., "Hepatic Subcellular Localization of CBDP-sensitive Soman Binding Sites," *Biochem. Pharmacol.*, Vol 40, pp. 1733-1737, 1990.
10. Pelham, H.R.B., "The Retention Signal for Soluble Proteins of the Endoplasmic Reticulum," *Trend. Biochem. Sci.*, Vol 15, pp. 483-486, 1990.
11. Gentry, M.K. and Doctor, B.P., "Amino Acid Alignment of Cholinesterases, Esterases, Lipases, and Related Proteins," In: *Enzymes of the Cholinesterase Family*, Quinn, D.M., Balasubramian, A.S., Doctor, B.P. and Taylor, P., eds., Plenum Press, New York, pp. 493-505, 1995.

Measuring Visual Resolution in the Contrast Domain: The Small Letter Contrast Test

LTC Jeff Rabin*

U.S. Army Aeromedical Research Laboratory
Fort Rucker, AL 36362-0577

Abstract

Previous studies showed that small letter contrast sensitivity (CS) is more sensitive than visual acuity (VA) to blur, light intensity, binocular vision, and visual differences among pilot trainees. It would be valuable to make this test available for general use. We developed a hard copy (letter chart) version called the Small Letter Contrast Test (SLCT) and evaluated its sensitivity and reliability in comparison to standard vision tests. The SLCT has 14 lines of letters with 10 letters per line. The letters are of constant size (20/25 at 4m), but vary in contrast by line in 0.1 log steps (0.01 log units per letter). Normal room illumination is used. The SLCT was evaluated in 16 subjects under various conditions (i.e., spherical and astigmatic blur, low luminance, one eye vs. two) to determine test sensitivity and reliability, and in patients with clinical conditions. Scores were compared to those obtained with standard tests of VA and CS. SLCT scores were similar to previous studies, and retest reliability was one line. The SLCT was more sensitive than VA to blur, luminance, and vision with two eyes vs. one. Greater sensitivity of the SLCT endured despite correction for variability. The SLCT was more sensitive than standard tests to visual loss from early cataract, keratoconus, corneal infiltrates, edema, and amblyopia. The SLCT reveals subtle visual loss that escapes detection by standard techniques. It will prove useful for monitoring vision in refractive surgery, corneal and macular edema, optic neuritis, and for evaluation of candidates for occupations requiring unique visual abilities like space and aviation.

1. Introduction

Optimal visual acuity (VA) is the goal of clinical vision care. Refraction of the eye, detection, diagnosis and treatment of ocular disease, and refractive surgery share the common goal of achieving best VA. The effectiveness of this approach is based on the fact that VA is a sensitive index of blur. Blur causes a reduction in VA, and this reduction is proportional to the amount of blur. But blurring the retinal image also reduces the contrast of fine detail (higher spatial frequencies). Due to the steep slope of the spatial contrast sensitivity (CS) function near the acuity limit, a loss of VA is associated with a greater loss of CS for higher spatial frequencies.

This principle is illustrated in Figure 1 which shows that 1D of defocus shifts the descending limb of the CS function downward and to the left. The shift leftward along the spatial frequency axis represents the loss of VA. The shift downward along the contrast dimension demonstrates the greater loss of CS for higher spatial frequencies.

The greater loss of CS than VA suggests that small letter CS may provide a more sensitive index of blur. In previous studies, we used letters displayed on a computer monitor to show that small letter CS is more sensitive than VA to small amounts of blur, subtle changes in stimulus luminance, vision with two eyes compared to one, and for identifying visual differences among pilot trainees (Rabin, 1994a, 1994b; 1995a, 1995b, 1995c; 1996). To make this test available for general use, we developed a hard copy (letter chart) version called the Small Letter Contrast Test (SLCT). This paper describes the design of the SLCT, its reliability, and its sensitivity for detecting visual differences from normal, as well as changes in vision over time.

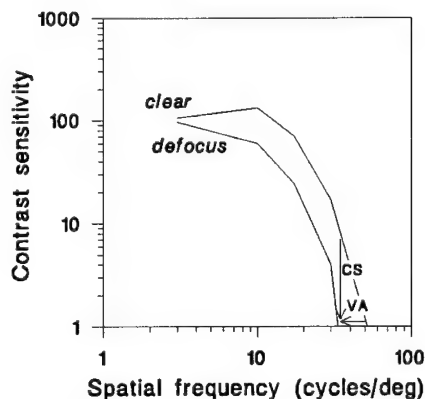


Figure 1. Defocus effects on visual acuity and high spatial frequency contrast sensitivity.

2. Methods

The SLCT is generated from computer software (Adobe Photoshop version 2.0.1), printed from a Kodak XL 7700 continuous tone digital printer, and heat-pressed to gator board (45 x 30 cm; fig. 2). It has 14 lines of letters with 10 letters per line. Unlike VA charts (e.g., Bailey and Lovie, 1976), which have letters that vary in size, the SLCT has letters of constant size (20/25 at 4 m) which vary in contrast, by line, in 0.1 log unit steps. Credit is given for each letter read correctly (0.01 log units per letter). Normal room illumination is used (fluorescent or incandescent), and letter contrast remains constant across a range of photopic room luminances (50 to 200 cd/m²).

To evaluate the sensitivity and reliability of the SLCT, performance on this new test was compared to that obtained with standard tests of visual function. High and low contrast VA (Bailey-Lovie acuity charts; Bailey and Lovie, 1976), large letter CS (Pelli-Robson chart; Pelli et al., 1988), and small letter CS (SLCT) were measured monocularly in 21 subjects with normal vision (ages 23 to 60 years) to establish means and confidence intervals. More extensive measurements were obtained from 16 of the 21 subjects (ages 23-35) to assess the sensitivity and reliability of each test to several conditions in which vision was compromised by small amounts. Subjects were tested under the following conditions in this order: (1) monocularly (right eye) with best optical correction, (2) low degree of spherical blur (+0.5 D) to simulate low myopia which would preclude an officer candidate from being accepted for pilot training, (3) low degree of astigmatic blur (+1 D x 90), also a failure criterion for pilot training, (4) low level of luminance (6 cd/m²) to represent the luminance of a night vision goggle display, (5) binocularly to compare to monocular scores, and (6) again monocularly with best correction to assess retest reliability.

Each subject initially was refracted to best VA and then tested in a vision laboratory illuminated by overhead fluorescent lighting under rheostat control (chart luminance=100 cd/m²). The subject was seated comfortably and wore a trial frame such that optical correction, different power lenses, and filter (low luminance condition) could be placed before the eye. Test distance for high and low contrast VA and SLCT was 4 m, while the Pelli-Robson chart was viewed at 1 m, as specified by the manufacturer. Each chart has two versions, which made it possible to change letter sequence between trials to discourage learning effects. In addition to subjects with normal vision, several patients with subtle visual loss from various conditions were evaluated. Informed consent was obtained from all subjects after protocol review by our institutional review committees.

3. Results

3.1 Normative Data

A clinical vision test can be used to determine if vision *differs from normal* or *changes over time*. These are separate issues requiring distinct statistical comparisons. For example, a patient presents with a history of refractive surgery, and testing is conducted to determine if vision is within normal limits. Vision is below normal if scores fall below the 95% confidence limit for normals (Table 1; *Detecting differences from normal*). To determine if the same patient's vision changes over time, it is necessary to obtain at least two measures from the patient, separated in time, to determine if the change in vision falls within the 95% confidence interval for change in normals. This interval, known as the coefficient of repeatability (Bland and Altman, 1986), is shown in log units and as lines of letters for each test (Table 1; *Detecting change over time*). A loss of one line of letters represents a significant decrease in vision on each test.

Row		LogCS
1	D R E Z U F V N H P	0.0
2	R P F D U N Z E V H	0.1
3	H R P E D V Z N F U	0.2
4	Z E D H P U V R N F	0.3
5	H P F D N V R E V H	0.4
6	R Z F D P N U E V H	0.5
7	V D H P E N R H U Z	0.6
8	U R N E D Z H F V P	0.7
9	N V Z F H E P R D U	0.8
10	D V N Z R H F U P E	0.9
11	P H V D F U E Z N R	1.0
12	R V U N D P H Z E F	1.1
13	F R E U P Z H D V R	1.2
14	E R P D N Z F U V H	1.3

Figure 2. The Small Letter Contrast Test (letter spacing and contrast levels not to scale).

Table 1. Detecting Differences from Normal and Change Over Time

Vision Test	Detecting differences from normal*		Detecting change over time	
	Mean \pm 2SD† (mean Snellen VA)	Below normal	Repeatability (log units)‡	Significant change (lines on chart)§
High contrast VA	-0.11 \pm 0.12 (20/16)	20/21 or less	0.06	4/5 line
Low contrast VA	0.00 \pm 0.14 (20/20)	20/28 or less	0.07	4/5 line
Pelli-Robson CS	1.88 \pm 0.17	1.70 or less	0.12	1 line
SLCT	1.21 \pm 0.18	1.02 or less	0.11	1 line

*Normals tested monocularly (n=21; 23 to 60 years). †Decimal units are logMAR for VA and logCS for PRCS and SLCT.

‡Coefficient of repeatability=2xSD of within-subject differences between scores. §1 line=0.1 log unit for VA & SLCT; 0.15 log units for PRCS.

3.2 Test Sensitivity

Figure 3 shows results from normal subjects for spherical blur, astigmatic blur, low luminance, and one eye vs. two eyes. The mean (\pm 1SE) reduction in vision is plotted in log units for each vision test. The reduction in vision was computed by taking the difference between log scores under normal and test conditions. Figure 3 shows that 0.5 D of spherical blur reduced high and low contrast VA by only 0.1 log unit (one line of letters), but produced a larger, 0.3 log unit (3-line) decrease on the SLCT. Little change was observed with the Pelli-Robson chart which uses large letters, and is thus unaffected by small amounts of blur.

A similar, albeit larger, effect was observed with astigmatic blur (+1 D x 90; fig. 3). There was a 2-line reduction in high and low contrast VA, but a greater 5.5-line reduction on the SLCT. While defocus simulates refractive error, a decrease in chart luminance can reduce vision, perhaps in a manner similar to that imposed by media opacities. Figure 3 shows that reducing luminance (from 100 to 6 cd/m²) produced a 1-line decrease in high contrast VA, a 2-line decrease in low contrast VA, a modest reduction of 1.3 lines on the Pelli-Robson chart, but a greater 5-line decrease on the SLCT. As in previous research (Campbell and Green, 1965), vision with two eyes compared to one produced only a slight improvement in high and low contrast VA (two letters; fig. 3), but a greater 1.3-line improvement in CS.

These results suggest that the SLCT is more sensitive than standard tests to blur, low luminance, and binocular enhancement. However, a larger effect does not ensure greater test sensitivity if variability is also greater. To standardize scores with respect to variability, the difference between each score and the value under optimal conditions was divided by the standard deviation of the measurement. This transformation, which expresses all test scores in common units of standard deviations, allows for direct comparison between results of different tests. Figure 4 shows test sensitivity standardized relative to variability between (top) and within subjects (bottom). Even when this correction is made for variability, the Small Letter Contrast Test still proved to be at least 2x more sensitive than standard tests under most conditions.

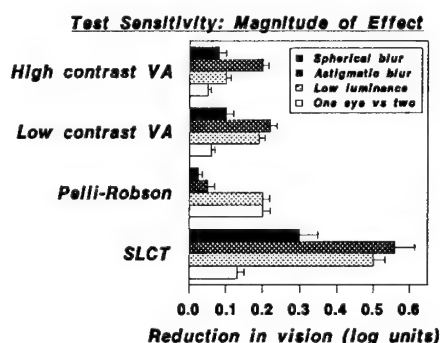


Figure 3. Mean magnitude of the effect for each condition is plotted for each test.

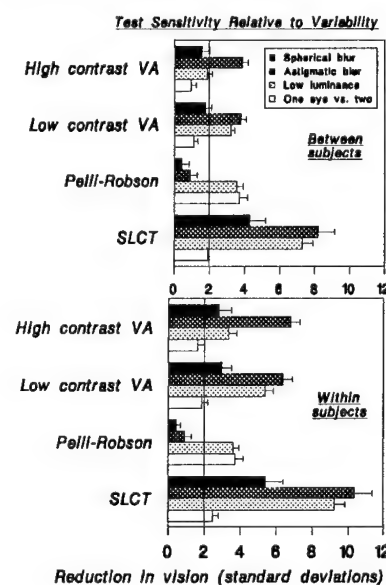


Figure 4. Test sensitivity (standard deviations) relative to variability between (top) and within subjects (bottom).

3.3 Application In Clinical Conditions

Figure 5 shows results of letter chart testing in several clinical conditions characterized by subtle visual loss. For each subject, the reduction in vision is plotted as standard deviations below the mean for normals ($n=21$). In each condition, including early nuclear cataract, previously undiagnosed keratoconus, mild amblyopia, corneal infiltrates, and contact lens edema, the SLCT is clearly as sensitive, if not significantly more sensitive than standard vision tests for detecting subtle differences from normal.

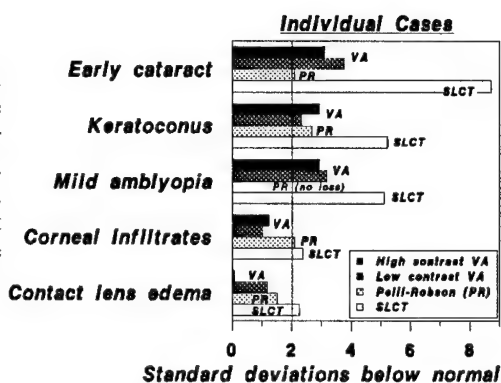


Figure 5. Test sensitivity in cases of visual loss.

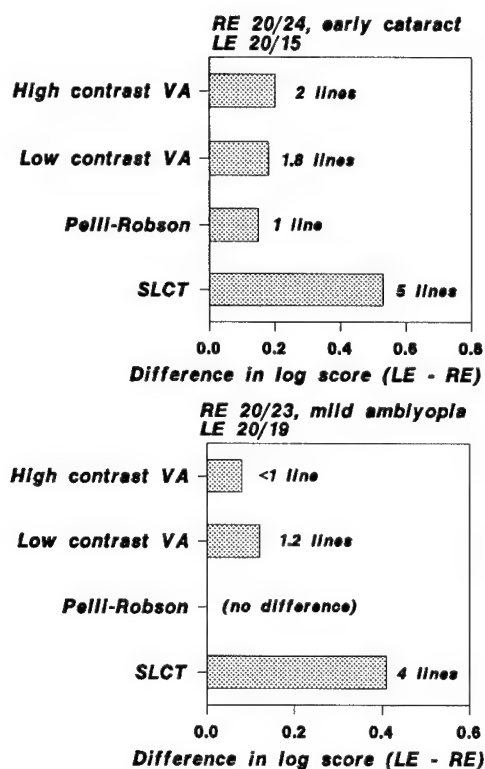


Figure 6. Two cases of subtle visual loss in one eye. The difference between log scores (good eye - affected eye) is plotted for a patient with early cataract (top) and for a patient with mild amblyopia (bottom).

Figure 7 shows results for a senior military aviator. The patient previously wore glasses, but recently switched to a monocular contact lens correction. He wears no lens in the right eye, but a soft toric lens in the left eye to correct the greater degree of astigmatism. While the patient is pleased with his contact lens correction, and achieves VA slightly better than 20/20, there is still an improvement with best spectacle correction compared to the habitual correction: no lens on right eye, soft lens on left eye. The improvement in vision, expressed as lines of letters on each chart, is plotted for each vision test. While best correction afforded only a 1-line improvement in high contrast VA, there was a 2-3 line improvement on the SLCT. Such exacting information may be used to guide decisions regarding the use of contact lenses in flight.

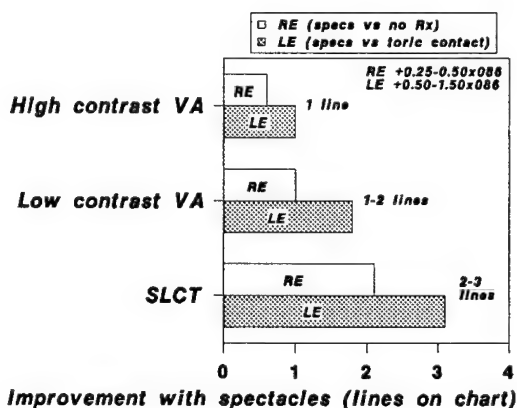


Figure 7. The improvement in vision with full spectacle correction as compared to the patient's habitual correction: no lens on right eye, soft toric contact lens on left eye.

4. Discussion

Like other letter chart tests, the SLCT is easy to give and easy to score. But despite its simplicity, the SLCT offers an extremely sensitive approach for monitoring visual function. It uses a forced-choice letter recognition task that is unaffected by shifts in patient criterion. Invaluable design principles of Bailey and Lovie (1976) and Pelli et al. (1988) are included and enhanced in the SLCT. By using a step size of 0.1 log units and 10 letters per line, a scoring precision of 0.01 log units per letter is achieved, which is 2x finer than existing tests. The unique feature of the SLCT, the use of small letters to measure CS, exploits the steep slope of the CS function where small changes in VA are associated with large changes in CS.

The present results confirm previous findings that small letter CS is more sensitive than VA to small amounts of blur, modest changes in stimulus luminance, and vision with two eyes compared to one (Rabin, 1994-1996). These findings, initially demonstrated with letters generated on a computer monitor, were shown to be valid for the SLCT--a letter chart now available for general use. Previous results for spherical blur were confirmed and a similar effect was found for astigmatic blur. A modest reduction in luminance produced a small decrease in VA, but a larger decrease on the SLCT. This effect on CS has been attributed to the quantal nature of light for which decreases in intensity are not matched by proportional decreases in noise (Rabin, 1994b). Visual enhancement with two eyes compared to one was greater for CS than VA, a result also related to the steep slope of the CS function (Campbell and Green, 1965). Most importantly, all results were validated statistically since greater sensitivity of small letter CS endured despite correction for variability.

Several clinical conditions, including early cataract, keratoconus, and mild amblyopia, were characterized by subtle losses of VA, but larger decreases on the SLCT. Since both VA and the SLCT use small letters composed of high spatial frequencies, and given the steep slope of the CS function, any condition that reduces VA should have a greater effect on the SLCT. The clinical cases described here exemplify this principle and underscore the sensitivity of the SLCT for detecting subtle visual loss. The SLCT should prove useful when there is loss of central vision undisclosed by standard tests. Applications include monitoring the course of vision in refractive surgery, macular edema, optic neuritis, cataracts, and evaluation of individuals for unique occupations such as space and aviation.

References

1. Rabin, J., "Optical defocus: Differential effects on size and contrast letter recognition thresholds," *Invest. Ophthalmol. Vis. Sci.*, Vol 35, No. 2, pp. 646-648, February 1994.
2. Rabin, J., "Luminance effects on visual acuity and small letter contrast sensitivity," *Optom. Vis. Sci.*, Vol 71, No. 11, pp. 685-688, November 1994.
3. Rabin, J., "Two eyes are better than one: Binocular enhancement in the contrast domain," *Ophthalm. Physiol. Opt.*, Vol 15, No. 1, pp. 45-48, January 1995.
4. Rabin, J., "Small letter contrast sensitivity: An alternative measure of visual resolution for aviation candidates," *Aviat. Space Env. Med.*, Vol 66, No. 1, pp. 56-58, January 1995.
5. Rabin, J., "Time-limited visual resolution in pilot trainees," *Military Medicine*, Vol 160, No. 6, pp. 279-283, June 1995.
6. Rabin, J., "Correction of subtle refractive error in aviators," *Aviat. Space Env. Med.*, Vol 67, No. 2, pp. 161-164, February 1996.
7. Bailey, I.L. and Lovie, J.E., "New design principles for visual acuity letter charts," *Am. J. Optom. Physiol. Opt.*, Vol 53, No. 11, pp. 740-745, November 1976.
8. Pelli, D.G., Robson, J.G. and Wilkins, A.J., "The design of a new letter chart for measuring contrast sensitivity," *Clin. Vis. Sci.*, Vol 2, pp. 187-199, 1988.
9. Bland, J.M. and Altman, D.G., "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, Vol 1, pp. 307-310, 1986.
10. Campbell, F.W. and Green, D.G., "Monocular versus binocular visual acuity," *Nature*, Vol 208, pp. 191-192, 1965.

Production of Recombinant Antibodies for Biosensor Applications

Peter A. Emanuel* and James J. Valdes

Research and Technology Directorate, U.S. Army Edgewood Research Development and Engineering Center, Aberdeen Proving Grounds, MD 21010.

Mohyee E. Eldefrawi

Department of Pharmacology and Experimental Therapeutics, School of Medicine, University of Maryland, 655 West Baltimore Street, Balt., MD 21201.

ABSTRACT

The need to isolate clones with rare binding specificities has required the use of a highly efficient method of screening and isolating antibody clones. The production of recombinant antibody libraries in bacteria allows for a more controlled selection of immunoglobulin specificity and can be used in circumstances where hybridoma fusions are unable to isolate rare clones with the desired epitope specificity. Initial screens of a combinatorial library directed against botulinum neurotoxin complex revealed clones which recognized the non neurotoxin proteins of the toxin complex similar to monoclonal antibodies produced by conventional hybridoma fusions. The combinatorial library was re-screened in order to isolate antibodies that specifically recognized the neurotoxin component of the toxin complex. The ability to alter the biopanning selection provides the researcher a measure of control over the selection process not available with traditional hybridoma fusions. The recombinant Fabs were biomanufactured in bacterial fermentation systems, purified through an engineered affinity tag and mounted on a fiber optic immunosensor to detect botulinum toxin.

1. INTRODUCTION

When performing a hybridoma fusion the probability of isolating antibodies to epitopes of an antigen which are not immunodominant is limited to a small portion of the $\approx 10^8$ immunoglobulins in the mouse repertoire. The need to isolate a larger proportion of the immune repertoire has prompted researchers to develop techniques for the construction of recombinant antibody libraries (Barbas, 1991; Posner, 1993). A cDNA library can be made from mouse messenger RNA isolated from the spleens of immunized mice. The cloned mouse heavy and light chain genes can be expressed in *Escherichia coli* to produce single chain or Fab antibody fragments, which are capable of antigen recognition (Better, 1988; Persson, 1991; Hogrefe, 1994). By expressing recombinant immunoglobulin repertoires in phage display libraries it is

possible to enrich and screen millions of prospective clones for antigen recognition (Garrard & Yang et al., 1991). The diagnostic and therapeutic uses of antibodies often require high affinities that can be difficult to obtain with hybridoma technology (Ward, 1992).

This study demonstrates the efficiency of a recombinant antibody approach using combinatorial phage display libraries in the production of mouse antibodies towards botulinum neurotoxin (NT) serotype B; a potent neurotoxin produced by the bacteria *Clostridium botulinum*.

2. RESULTS

Antibodies against botulinum toxin type B complex were raised in female BALB/C mice after immunization by sequential subcutaneous injections of human pentavalent (A, B, C, D, E,) toxoid emulsified in FCA. The polyclonal serum was shown to be responsive towards botulinum toxin complex serotypes A and B (>500 kDa), with an endpoint titer of 1:100,000, as well as towards pure botulinum NT B (150 kDa), that lacked the hemagglutinin commonly found associated with the toxin complex. From the ELISA data, it was concluded that the challenged mice had raised an immune response directed against the 150 kDa NT B as well as towards the non-NT proteins of the toxin complex.

To screen the hybridoma cell lines ELISA assays were performed using four preparations of botulinum toxin (complexes of type A and B and the corresponding 150 kDa NTs) immobilized to microtiter wells. A hybridoma cell line was isolated which exhibited antigen specificity towards botulinum toxin complex A and B as demonstrated by ELISA (Table 1; monoclonal). This clone did not respond to the 150 kDa NT types A or B; indicating specificity towards non-NT proteins of the botulinum toxin complex.

Cell line name	Complex A (500 kDa)	Complex B (500 kDa)	NT A (150 kDa)	NT B (150 kDa)	BSA 66 kDa
Bot01 monoclonal	1.88	2.82	0.03	0.04	0.09
BotFab 1	1.85	2.7	0.03	0.03	0.33
BotFab 5	0.62	1.4	0.04	0.07	0.05
BotFab 7	2.08	2.8	0.05	0.05	0.05
BotFab 8	1.73	2.4	0.09	0.07	0.09
BotFab 12	1.35	2.05	0.04	0.05	0.06
ToxFab 146	0.03	0.03	0.02	1.41	0.03
ToxFab 149	0.04	0.03	0.02	1.06	0.02

Table 1: ELISA characterization of antibody specificity towards botulinum toxin. ELISA assays were performed by immobilizing each indicated botulinum toxin serotype on the bottom of microtiter wells. Monoclonal antibody or bacterially expressed Fab protein served as the primary antibody followed by goat anti-mouse Fab HRP conjugate (1:2500) as a secondary colorimetric probe. The results of the assay for selected clones are shown below as absorbance at 405 nm which have not been normalized for protein concentrations. Botulinum (Bot) monoclonals are affinity purified IgG from hybridoma cell lines. BotFab and ToxFab are diluted media supernatants from IPTG induced bacterial clones.

RNA isolated from the spleens of a second set of mice immunized similarly were used for the construction of a combinatorial Fab antibody

library. A primary recombinant antibody library of 3×10^6 pfu was constructed and excised as a double stranded phagemid (Hay and Short, 1992). Co-infection of the cells with the VCSM13 helper virus induced production of filamentous bacteriophage, that express the heavy and light chain polypeptides on the surface of the virus. The display of the Fab fragment on the surface of the phage allows for the enrichment of antigen specific clones through biopanning against botulinum type B complex immobilized in the wells of microtiter plates.

Biopanning enrichment and screening with ^{125}I -labeled botulinum type B complex revealed numerous positive clones. Forty eight identified by ELISA were selected and subcloned into the pHist purification vector. The Fab genes were expressed as fusions with a histidine hexapeptide tail, rather than as fusions with the cpIII protein. Commercially available Ni-NTA resins bind histidine oligomers with high affinity to allow for a one step purification of the Fab protein from crude bacterial lysates. Antibody protein was eluted under non reducing conditions with 500 mM imidazole. SDS-PAGE analysis of the purified Fab polypeptide shows a 50 kDa protein under non reducing conditions and two polypeptides of 31 and 29 kDa under reducing conditions (Figure 1; lanes 4 and 5). A western blot of SDS-PAGE gels, run under non reducing conditions and probed with mouse anti-Fab antiserum, confirmed that the 50 kDa polypeptide was mouse Fab. Western blot of the purified Fab protein under reducing conditions using mouse anti-kappa light chain antiserum confirmed that the light chain was the 31 kDa polypeptide (data not shown).

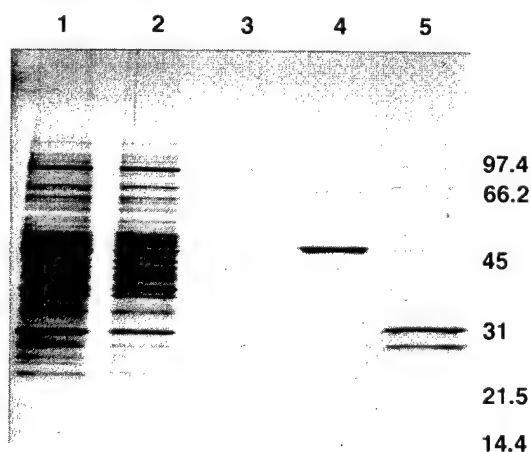


Figure 1: Silver stained SDS-PAGE gel of Fab antibody preparations.

Lane 1: 5 ul crude bacterial lysate, Lane 2: 5 ul Ni-NTA column unbound fraction, Lane 3: 5 ul Column wash, Lane 4: 7 ul of purified Fab protein under non-reducing conditions, Lane 5: 7 ul of purified Fab protein under reducing conditions. Molecular weight standards are along the right side of the gel.

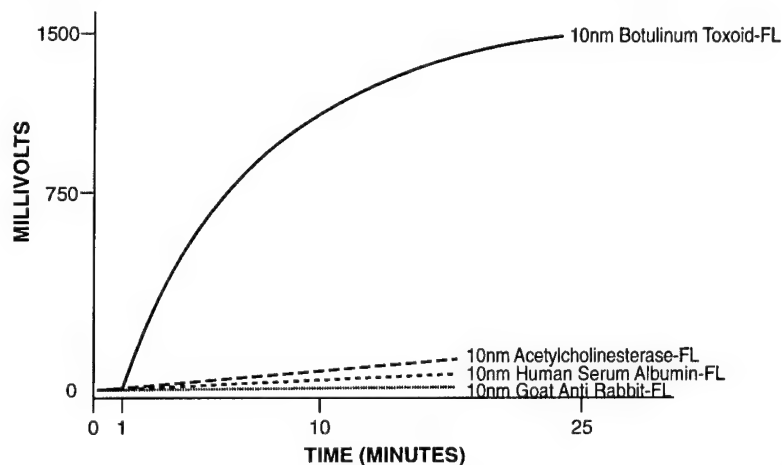
The results demonstrate that the bacterially expressed mouse

Fab could be isolated in a highly purified form following metal chelate affinity chromatography, at levels of approximately 1600 ug/liter of LB broth. Using these techniques 65 clones were isolated, twenty seven were positive by ELISA for botulinum types A and B toxin complexes and none were reactive to the 150 kDa NT types A and B, that lacked the non-NT proteins found in the complex. The ELISA results of several representative Fab clones are presented in Table 1.

To direct the antigen selection towards the NT a frozen aliquot of the primary unenriched library was thawed, expanded and biopanned against 150 kDa type B NT. This resulted in the selective enrichment of clones capable of recognizing only the 150 kDa NT component of the botulinum toxin complex.

Two Fab clones were isolated which specifically recognized the 150 kDa botulinum B neurotoxin component in an ELISA assay, but did not react with either the immunologically distinct purified NT type A or the botulinum A and B toxin complexes (Table 1; ToxFab clones). These results demonstrated that a diversity of epitopic responses resided in the primary antibody library and that clones, which produce Fabs of a given specificity, can be selected from the primary antibody library by a directed biopanning procedure.

Figure 2: The recombinant Fabs were then covalently immobilized to a quartz fiber at 100 ug/ml and placed in the sensing chamber of a fiber optic fluoro-immunosensor. The detection



of this device is based on the excitation of a fluorescein linked to the botulinum toxoid which binds to the antibody immobilized to the surface of a quartz optic fiber. A signal is generated upon binding of the fluorescein labeled botulinum toxoid (Bot Tox-FL) by the evanescent wave associated with the light traveling within the fiber. Bot Tox-FL in the bulk flow solution, beyond the evanescent field, does not contribute to the fluorescence transmitted by the fiber. In order to demonstrate the specificity of the immunosensor the ability of the Fab coated fibers to bind Bot Tox-FL and three other fluorescein labeled non-target proteins (acetylcholinesterase-FL, goat anti-rabbit IgG-FL, and human serum albumin-FL) was investigated. The time course of binding of the different fluorescent ligands, all tested at 10nM in PBS/casein, indicated that none of the three non-target fluorescent proteins bound to the fiber whereas Bot Tox-FL bound well. The nonspecific binding is approximately 2% of the Bot. Tox.-FL.

3. DISCUSSION

Recombinant antibodies produced in bacterial systems can be used to overcome limitations associated with traditional hybridoma technology. Using botulinum NT as an example, we show that isolation of antibodies recognizing unique epitopes can be achieved with combinatorial libraries.

The re-selection of the antibody library demonstrates that conditions for biopanning enrichment can be altered to select for different antigens of a multicomponent antigen complex. The stringency of washing during the process can be changed in order to isolate antibodies with high or low affinity (Garrard & Yang et al., 1991). Alterations in the biopanning gives an additional level of control which allows the selection process to be driven towards a particular antigen or epitope. The production of antibody "superlibraries" is also possible. These large cDNA libraries would enable the Army to bio-pan a new biothreat agent and select for antibodies which recognize it. The Fab fragments could be isolated in a cloned form within days of initial exposure and scaled up within 2 weeks for use on a BIDS system platform.

The technique offers, in addition to the ability to manipulate the selection process, other attributes over traditional hybridoma technology. Antibody producing clones are available in *E. coli* bacteria as a double stranded phagemid. The genetic manipulation of this plasmid DNA is straightforward and provides the capability to alter the affinities by mutagenesis or by chain shuffling (Marks et al., 1993; Jespers, Roberts et al., 1994). Through the use of convenient restriction sites the Fab gene insert can be subcloned into a new vector, which allows the Fab to be expressed as a fusion with a histidine hexapeptide tag on the carboxyl terminus.

The need for antibodies with particular specificities and affinities has often encountered difficulties using standard technologies. Bacterial expression of Fab fragments is superior to hybridoma production because of the ease, rapidity and reduced cost of biomanufacturing via a fermentation facility. The power of phage display lies in the vast numbers of clones that can be screened and enriched in order to isolate unique specificities. As this technique gains wider acceptance for use in isolating immunoglobulins significant modifications may be introduced to extend its utility beyond immunodetection for a dual use in therapeutic applications

4. REFERENCES

- Amberg, J., H. H. Hogrefe, et al. (1993). SurfZAP Vector: Linking Phenotype to Genotype for Phagemid Display Libraries. *Strategies in molecular biology* 5, 2-5.
- Amberg, J., H. Hogrefe, et al. (1993). SurfZAP vector: Linking phenotype to genotype for phagemid display libraries. *Strategies in molecular biology* 6, (1), 2-4.
- Barbas, C., A. Kang, R. Lerner and S. Benkovic (1991). Assembly of combinatorial antibody libraries of phage surfaces: the gene III Site. *Proc. Natl. Acad. Sci* 88, 7978-7982.
- Chomczynski, P. and N. Sacchi (1987). Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* 162, 156-159.
- Hay, B. and J. M. Short (1992). ExAssist helper phage and SOLR cells for lambda ZAP II excisions. *Strategies in molecular biology* 5, (1), 16-18.
- Hogrefe, H. and Bob. Shopes. (1994). Construction of Phagemid Display Libraries with PCR-amplified Immunoglobulin Sequences. *PCR Methods and Applications* S109-S122.
- Jespers, L., A. Roberts, et al. (1994). Guiding selection of human antibodies from phage display repertoires to a single epitope of an antigen. *Bio/Technology* 12, 899-903.
- Marks, J., W. Ouwehand, J. Bye, R. Finnern, B. Gorick, D. Voak, S. Thorpe, N. Jones and G. Winter (1993). Human Antibody Fragments Specific for Human Blood Group Antigens from a Phage Display Library. *Bio/Technology* 11, 1145-1149.
- Posner, B., I. Lee, T. Itoh, J. Pyati, R. Graff, G. Thorton, R. LaPolla and S.J. Benkovic (1993). A revised strategy for cloning antibody gene fragments in bacteria. *Gene* 128, 111-117.
- Ward, E. S. (1992). Antibody Engineering: The use of *Escherichia coli* as an expression host. *FASEB J.* 6, 2242-2427.

A Protein Engineering Approach to Designing Staphylococcal Enterotoxin Vaccines

Robert G. Ulrich*

**U. S. Army Medical Research Institute of Infectious Diseases
Fort Detrick, Frederick, MD 21702-5011**

Abstract

A vaccine was designed by site-directed mutagenesis of staphylococcal enterotoxin B protein surfaces in contact with receptors. The mutagenesis strategy was guided by molecular modelling of receptor and ligand complexes. To identify the important molecular features of toxicity, the interactions of staphylococcal enterotoxin B with two essential cellular receptors were first addressed with mutant toxin that had diminished binding to either the major histocompatibility complex class II molecule or the T-cell antigen receptor. Attenuation of several disease markers and retention of antigenic structure were evaluated for each engineered mutation. The greatest disruption of biological activity was obtained by targeting interactions with the major histocompatibility class II receptor. The design of the vaccine was then optimized by computational studies of the effects of replacing essential amino acid residues. The final engineered product was found to be nontoxic and highly immunogenic in mice and rhesus monkeys.

1. Introduction

Staphylococcal enterotoxin B (SEB) is an exotoxin produced by certain strains of *Staphylococcus aureus*. At least 7 known staphylococcal enterotoxins, toxic shock syndrome toxin-1 (TSST-1) and the streptococcal pyrogenic exotoxins A-C are commonly referred to as bacterial superantigens because their pathological effects are exerted through the immune system. A hallmark of these soluble 23-29 kDa bacterial protein toxins is an affinity for two receptors that control immune responses: the major histocompatibility complex (MHC) class II molecule and T-cell antigen receptors (TCRs). By stimulating a substantial number of T cells, the complex of MHC class II molecule, toxin and TCR initiates a release of toxic levels of cytokines. The symptoms produced by each toxin are similar and are characterized by a rapid drop in blood pressure, elevated temperature, rash followed by a desquamation of the palms of the hands and multiple organ failure.

Exposure to native SEB inactivates critical components of the immune response that are required for antibody production. By binding to the toxin when it enters the body, antibodies are likely to neutralize the pathological effects in vaccinated individuals. The conventional vaccine approach for protein toxins is to produce a formaldehyde-inactivated toxoid. However, formaldehyde treatment can potentially alter the antigenic structure of vaccines, and residual traces of this agent are also toxic. I have examined protein engineering as an alternative method to develop vaccines. To accomplish this minimally requires molecular cloning of the gene encoding the antigen, a method for expression of sufficient quantities of antigen, and the development of appropriate assays to determine toxic and immune potencies of the product. This effort is greatly aided by the availability of structural data for both antigen and receptor. The methods used in this study have been described in detail previously, and are cited where appropriate.

2. Results and discussion

2.1 Molecular modelling

Extensive structural and biochemical studies of both receptors and toxins were initiated to better understand the molecular nature of SEB binding. Homology modelling and x-ray crystallography studies have confirmed that most of the bacterial superantigens are structurally related (reviewed in Ulrich *et al.*, 1995a). Although SEB and TSST-1 appear to have little amino acid sequence in common, they exhibit homologous protein folds that are composed largely of β strands within two distinct domains (Fig. 1) and bind to receptors in a similar, but not identical, manner. We have found substantial antibody cross-reactions between each distinct toxin, most likely because of this conservation of protein structure.

The toxin amino acid residues in contact with the TCR are located in regions of high sequence variability, presenting a unique surface for interaction for each toxin. Because these receptor-binding sites are variable, different toxins stimulate unique subpopulations of T cells. However, common structural strategies are employed by these toxins (Ulrich *et al.*, 1995b) for



Figure 1. Staphylococcal enterotoxin B (left) and toxic shock syndrome toxin-1 (right) bound to the HLA-DR1 receptor (dark tracing). From the x-ray crystallography data of Jardetzky *et al.* (1994) and Kim *et al.* (1994).

binding to the α subunit of the human MHC class II molecules, HLA-DR (Fig. 1). A hydrophobic binding loop, interacting with the DR α subunit, is conserved in all superantigens, while a second binding site, consisting of a dominant polar pocket, is found in all except TSST-1. Other variable contacts, such as with the disulfide loop of SEB, serve to further stabilize receptor-toxin complexes.

The binding interface between SEB and HLA-DR1 (Jardetzky *et al.*, 1994) consists principally of two structurally conserved surfaces located in the N-terminal domain: a polar binding pocket derived from three β -strand elements of the β -barrel domain and a hydrophobic reverse turn. The binding pocket of SEB contains residues E67, Y89 and Y115, and binds K39 of the DR α subunit. The hydrophobic region of the binding interface between SEB and the HLA-DR1 molecule consists of SEB residues 44-47 centered around L45 and located in a large reverse turn connecting β -strands 1 and 2 of SEB. These residues appear to make strong electrostatic interactions with DR α through their backbone atoms (Jardetzky *et al.*, 1994).

2.2 Site-directed mutagenesis

The gene encoding the SEB molecule was cloned from a toxigenic strain of *S. aureus* and expressed as a recombinant protein in *E. coli* (Ulrich *et al.*, 1995b). Preliminary molecular simulation predictions suggested that receptor binding could be disrupted, without significant loss of tertiary protein structure. Attenuation of the biological activity of SEB was first addressed by generating site-specific mutants that resulted in diminished binding to either the HLA-DR1 molecule or the TCR (Ulrich *et al.*, 1995b).

2.3 Biological activity of SEB mutants in an animal model

The biological effects of the genetically-attenuated toxins were examined in mice. Unlike primates, the mouse MHC class II α -chain does not have a critical lysine (K39 of DR α) required for high-affinity binding. However, when a potentiating signal like lipopolysaccharide from Gram-negative bacteria (LPS) is present (Stiles *et al.*, 1993), the mouse toxicities are close to those found in primates. This observation provided us with a way to study toxicity and

Table 1. Effect of a reduction in receptor binding on the vaccine potential and toxicity of staphylococcal enterotoxins¹

Vaccine mutation inhibiting binding to	mouse lethality ²	antibody titer ³	Immune protection ⁴
T-cell receptor	20%	>1:10,000	100%
Major histocompatibility class II receptor	0%	>1:10,000	100%
Control wild-type toxin	100%	NT ⁵	NT
Control (Adjuvant only)	0%	<1:50	0%

¹ Adapted from Bavari *et al.* (1996a). Summary data for single residue SEA mutants are presented. TCR

² Mice (nonimmunized; n=10/group) were injected i.p with 10 LD₅₀ (10 µg) of mutant or wild-type toxin and 3h later with 150 µg LPS.

³ Serum dilution resulting in optical density readings of four times above the negative control, consisting of no SEA or serum added to the wells.

⁴ Mice (n=10/group) were immunized with 3 doses, two weeks apart, of site-specific mutants plus adjuvant or adjuvant alone. Two weeks after the final injection mice were challenged with 10 LD₅₀ of wild-type SEA plus LPS.

⁵ Not tested.

immunogenicity of the vaccines (Table 1). The wild-type toxin and control mutants were 100% lethal to mice at a dose of 10 LD₅₀, when coadministered with LPS, while mutants that were defective in HLA-DR1 binding were not lethal (Bavari *et al.*, 1996a). In contrast, a low but significant level of toxicity was exhibited by mutants that were defective only in TCR binding. The engineered SEB that were defective in receptor binding elicited high antibody titers and the vaccinated mice survived a challenge with 10 LD₅₀ of wild-type SEB. Eliminating receptor binding (TCR or HLA-DR1) appeared to facilitate conventional antigen processing and presentation, while wild-type or control mutants still inhibited immune responses (Bavari and Ulrich, 1996).

2.4 Optimized vaccine design

It was clear from the previous experiments that SEB mutations affecting HLA-DR1 binding were potentially safer for developing vaccines. Therefore, additional structural changes were engineered into the SEB surface interfacing with HLA-DR1 (Fig. 2). Mutation of residue Y89 in SEB to alanine resulted in 100-fold reduction in HLA-DR1 binding (Ulrich *et al.*, 1995b). The substitution of alanine for Y89 in SEB eliminated the hydrogen bond with K39 of the HLA-DRα subunit and disrupted packing interactions with adjacent protein residues. Modeling of the SEB mutant Y92A predicted an increase in solvent-accessible surface area for Y115, allowing the formation of a hydrogen bond to the carboxylate group of E67 and thus disrupting key anchoring and recognition points for HLA-DR1. Substitution of SEB Y115 with alanine also resulted in 100-fold reduction of binding. The K39 side chain of HLA-DRα forms a strong ion-pair interaction with the SEB E67 carboxylate group and hydrogen bonds with the hydroxyl groups of Y89 and

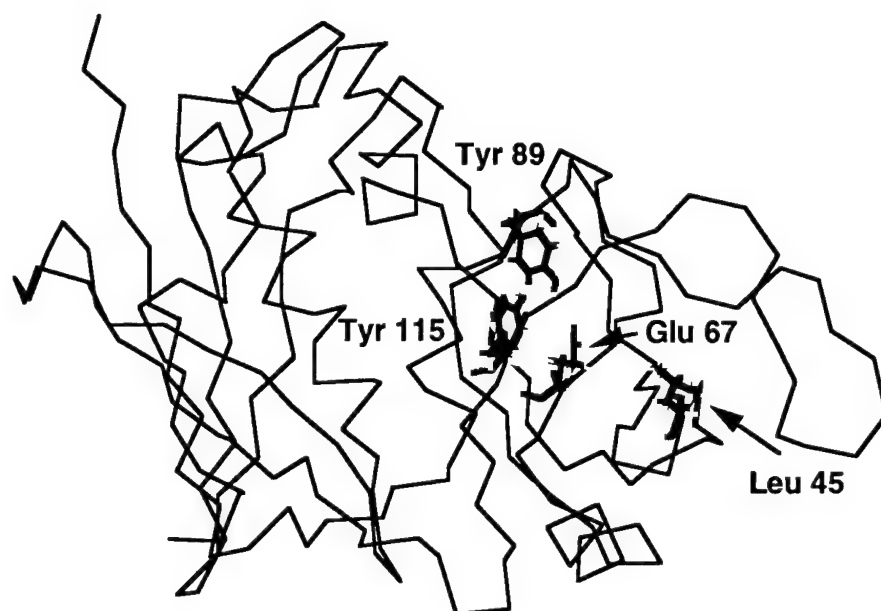


Figure 2. Key residues comprising the receptor-binding surface of staphylococcal enterotoxin B (Ulrich *et al.* 1995). The hydrophobic loop is centered around Leu 45, while Tyr 89, Glu 67 and Tyr 115 are hydrogen-bonded to a Lys 39 of the HLA-DR receptor.

Y115. Substitution of SEB E67 by glutamine reduced binding affinity by 100-fold, reflecting the replacement of the strong ionic bond with a weaker hydrogen bond. Within the hydrophobic binding loop, the mutation of L45 to an arginine reduced overall HLA-DR1 binding 10-fold, attributable to the less energetically favorable insertion of a highly charged residue into a hydrophobic depression on the DR1 molecule.

2.5 Vaccine efficacy in nonhuman primates

Based on the observations above, a final vaccine was produced that incorporated mutations of residues in both the polar pocket and hydrophobic loop binding motifs of SEB. This protein product exhibited no binding to either the mouse or human MHC class II receptors. Rhesus monkeys were injected with this vaccine have no detectable increase of serum cytokines and no apparent toxicity and high antibody titers were elicited in monkeys that received one immunization with the genetically attenuated vaccines (data not shown). The serological response of animals vaccinated with three doses of formalin-treated SEB toxoid (100 µg/injection) gave comparable results (Bavari *et al.*, 1996b), suggesting that the recombinant vaccine was very immunogenic. Immunized rhesus monkeys survived a lethal challenge with >10 LD₅₀ of wild-type SEB (Table 2). Collectively, these results suggest that the engineered SEB vaccine is safe, highly antigenic and effective at protecting the immunized individual from lethal aerosol exposure to SEB.

Table 2. Engineered staphylococcal enterotoxin B vaccine efficacy in rhesus monkeys

Treatment ¹	Antibody titer ²	Immune protection ³
Vaccine with adjuvant	>10,000	100%
Adjuvant only	<50	0%

¹Rhesus monkeys (n=10) were injected i.m. with 10 µg of SEB vaccine with Alhydrogel adjuvant. A total of 3 immunizations, 1 month apart were given. Controls (n=2) received only Alhydrogel.

²Serum dilution resulting in optical density readings of four times above the negative control, consisting of no SEB or serum added to the wells.

³Immunized and control rhesus monkeys were challenged with >10 LD₅₀ of wild-type staphylococcal enterotoxin B as an aerosol.

Acknowledgments

I thank my colleagues Sina Bavari, Mark Olson, Robert Hunt and Beverly Dyas for their many contributions to this project.

References

1. Ulrich, R.G., Bavari, S. and Olson, M., "Bacterial superantigen in human diseases: Structure, Function and Diversity," *Trends in Microbiol.* Vol 3, pp. 463-468, 1995.
2. Jardetzky, T. S., Brown, J. H., Gorga, J. C., Stern, L. J., Urban, R. G., Chi, Y., Stauffacher, C., Strominger, J. L. & Wiley, D. C., "Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen," *Nature*, Vol 368, pp. 711-718, 1994.
3. Kim, J., Urban, R. G., Strominger, J. L., & Wiley, D. C., "Toxic shock syndrome toxin-1 complexed with a class II major histocompatibility molecule HLA-DR1," *Science*, Vol. 266, pp. 1870-1874, 1994.
4. Ulrich, R.G., Bavari, S. and Olson, M., "Staphylococcal enterotoxins A and B share a common structural motif for binding class II major histocompatibility complex molecules," *Nature Struct. Biol.*, Vol. 2, pp. 554-560, 1995.
5. Stiles, B.G., Bavari, S., Krakauer, T. and Ulrich, R. G., "Toxicity of staphylococcal enterotoxins potentiated by lipopolysaccharide: major histocompatibility complex class II molecule dependency and cytokine release," *Infect. Immun.*, Vol. 61, pp. 5333-5338, 1993.
6. Bavari, S., Dyas, Beverly and Ulrich, R.G., "Superantigen Vaccines: A Comparative Study of Genetically Attenuated Receptor Binding Mutants of Staphylococcal Enterotoxin A," *J. Infect. Dis.*, in press, 1996.
7. Bavari, S., Olson, M. and Ulrich, R.G., 1996. "Bacterial superantigen vaccines," in: *Vaccines*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, NY. in press.

Molecular Beam Mass Spectrometric and Modeling Studies of Neat and NH_3 -Doped Low Pressure $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ Flames: Formation and Consumption of NO

R.C. Sausa* and W.R. Anderson
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

G. Singh
University of Maryland Eastern Shore
Princess Anne, MD 21853-1299

G.W. Lemire
Joint Contact Point Directorate
Dugway Proving Ground, UT 84022-5000

Abstract

An experimental and chemical modeling study of neat and NH_3 -doped $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ flames is conducted in order to understand the fundamental mechanism for NO conversion to the final products when trace concentrations of NH_3 are added. Temperature and species profiles are measured with a thin wire thermocouple and a quadrupole mass spectrometer, respectively. The experimental mole fractions are compared to both equilibrium and premixed, laminar flame code calculations. The flame code employs a chemical mechanism consisting of 83 reactions and 20 species with rate constants obtained from a critical literature review. Equilibrium calculations are in very good agreement with experimental measurements and flame code calculations for N_2O , N_2 , and H_2O in the post flame region of both neat and doped flames. However, they underpredict the H_2 and NO mole fraction in both neat and doped flames, proving that NO is not in equilibrium and prevents full energy release in the flames. The flame code profiles of the majority species agree very well with experiment for the neat flame and reasonably well for the doped flame. A 55% reduction in the NO mole fraction for 4% dopant is predicted in the post flame region, in good agreement with that observed experimentally. The flame calculations overpredict, however, the NH_3 mole fractions in the post flame region, suggesting that refinements in the model are necessary. Rate and sensitivity analyses performed on both neat and doped flames reveal that the decrease in NO mole fraction results from less NO formation in the doped flame by reactions $\text{N}_2\text{O} + \text{H} = \text{NO} + \text{NH}$ and $\text{N} + \text{OH} = \text{NO} + \text{H}$ and more of its consumption to N_2 by reactions $\text{NO} + \text{NH}_2 = \text{N}_2 + \text{H}_2\text{O}$ and $\text{NO} + \text{NH}_2 = \text{NNH} + \text{OH}$.

Introduction

An understanding of the way additives alter the chemical pathways in combustion systems is a prerequisite for controlling and enhancing system performance. In many systems NO is a product of incomplete combustion whose formation prevents full energy release of the system. It is also known to cause the formation of a dark zone in the burning of solid propellants which is undesirable, and as an environmental pollutant poses problems during demilitarization. Thus, it is of interest to investigate the incorporation of various additives in nitrogenous combustion systems for NO reduction. The $\text{H}_2/\text{N}_2\text{O}$ chemical system was chosen since it is fairly

simple and has important implications in understanding NO_x pollutant formation, and nitramine propellant combustion and decomposition. The elementary reactions of the system form subsets of larger mechanisms needed to understand the nitrogen chemistry of more complex combustion systems.

This paper reports combined experimental and modeling flame structure studies of neat and NH_3 -doped $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ flames in order to explore the role of the additive in converting NO to final products. NH_3 was selected as the additive because of its proven use in the thermal deNO_x process to remove NO from effluent streams of industrial furnaces. The experimental mole fractions are compared with both equilibrium and premixed laminar flame code calculations. Rate and sensitivity analyses of the flame code calculations reveal the mechanisms responsible for NO formation and consumption.

Experimental/Modeling

The details of the experimental apparatus and modeling used in this study are discussed elsewhere (Sausa et al., 1993 and Dayton et al., 1994). Briefly, a near stoichiometric, 30-Torr $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ flame was stabilized on a flat burner and doped with up to 7% NH_3 . Species concentration profiles of H_2 , N_2O , NH_3 , H_2O , N_2 , and NO were obtained by molecular beam mass spectrometry (MB/MS). The errors associated with the absolute species concentrations are $\pm 10\%$. The flame temperatures were measured with a Pt/Pt-Rh(10%) fine wire thermocouple coated with a beryllium oxide(15%)/ yttrium oxide mixture, and corrected for temperature radiation losses using OH laser-induced fluorescence. Doping the neat flame with 4% of NH_3 results in approximately a 2 mm displacement of the temperature profile away from the burner surface, and approximately a 44K drop in the peak temperature (1920K). The drop in peak temperature is consistent with NASA-Lewis equilibrium calculations which yield a difference in the adiabatic flame temperatures of only 9K. In the post flame region the measured flame temperatures are nearly identical, approximately 1822K at 13.75 mm. The overall uncertainty of temperature measurements is $\pm 100\text{K}$ in the region of the peak temperature.

The computations utilized the Sandia Laboratories flame code PREMIX (Ver.2.5b) which employs the CHEMKIN-II (Ver. 3.0) library. The calculations were performed with the experimental temperature profiles used as input to the flame code, and the inclusion of both thermal diffusion and multicomponent transport package option. The chemical mechanism used for the flame calculations consisted of 83 reactions and 20 species. The bulk of the reactions were obtained from a benchmark review of nitrogen chemistry in combustion (Miller and Bowman, 1989) which was updated for the $\text{H}_2/\text{N}_2\text{O}$ chemical system (Sausa et al., 1983 and Dayton et al., 1994). More recent updates include the rate constant for $\text{N}_2\text{O} + \text{M} = \text{N}_2 + \text{O} + \text{M}$ (Allen et al. 1994) and the efficiency factor for $\text{M} = \text{H}_2\text{O}$ (Glarborg et al., 1995). Finally, many recent results of Hanson and coworkers for reactions involving NH_x species have also been incorporated in the mechanism (Mertens et al., 1991).

Results/Discussion

Experimental and predicted mole fractions of H_2 , N_2 , H_2O , and N_2O for the neat $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ flame are presented in Figure 1A. The computed profiles accurately model the experimental results throughout the entire flame for the neat flame. Quantitatively, the calculated mole fractions agree well with those measured experimentally, particularly in the post flame region. The discrepancy in the N_2O profile near the burner surface may be in part related to the flame perturbation caused by the quartz sampler in that region. NASA-Lewis calculations performed using a temperature of 1822K yield equilibrium mole fraction values of 0.351 for N_2 and H_2O , and 0.0 for N_2O which are also in good agreement with the experimental mole fractions at 13.75 mm. The calculated H_2 equilibrium value of .0313 is, however, a factor of two smaller than the experimental value, indicating incomplete reaction.

Presented in Figure 1B are the experimental and calculated mole fraction profiles of the major species

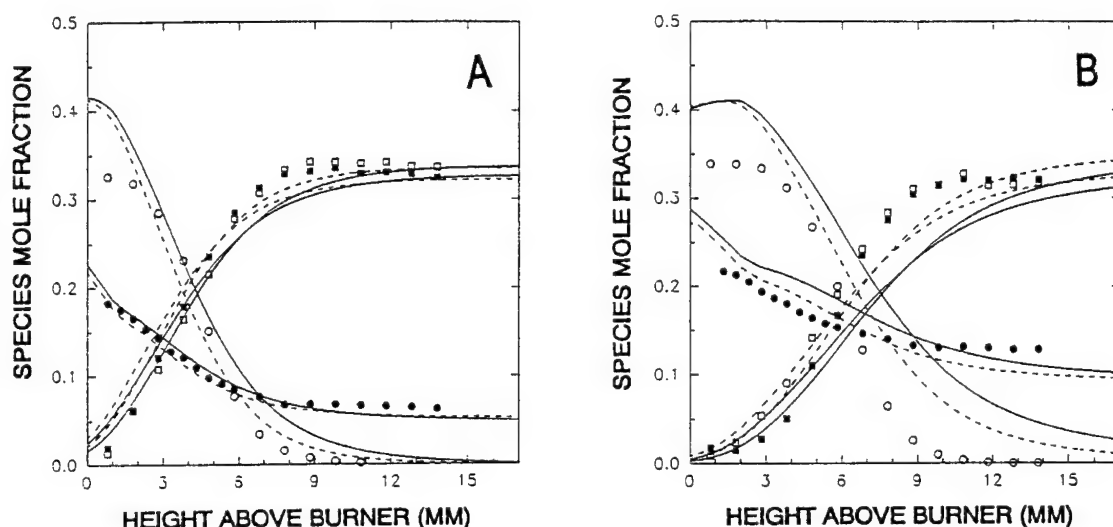


Figure 1. Calculated (—) and experimental H₂ (●), N₂O (○), H₂O (□), and N₂ (■) species profiles for neat (A) and 4% NH₃ doped H₂/N₂O/Ar flame (B). The dashed curves are calculated profiles generated by increasing the temperature profile by a factor of 1.05, the uncertainty in the experiment.

for the NH₃-doped flame. Species profiles generated by increasing the experimental temperature profile by 1.05, the experimental uncertainty, are also presented. Both the experimental and modeled profiles are shifted away from the burner compared to their positions in the neat flame. The result follows the trend in temperature profiles. For the doped flame, the modeled profiles are shifted approximately 2 mm further away from the burner surface compared to those measured, indicating that the overall chemical reaction rates in the model are too slow. The experimental and calculated mole fraction values for H₂, H₂O, and N₂ agree reasonably well, while the N₂O profile is overpredicted by the model. As in the neat flame, equilibrium mole fraction values for N₂, H₂O, and N₂O are in good agreement with the experimental values at 13.75 mm. The equilibrium H₂ mole fraction is, however, a factor of 1.4 lower than the experimental value.

The calculated and experimental NO mole fraction profiles for both the neat and NH₃-doped flames were also determined. Quantitatively, excellent agreement between model and experiment is obtained in the post flame region. The addition of 4% NH₃ reduces the NO concentration by approximately 45%, an effect well-predicted by the model. Also, a leveling off of the reduction efficiency is observed for NH₃ concentrations greater than 6%. A 5% change in the temperature profile inputted to the model results in approximately a ± 15 and 20 % change in the calculated NO mole fraction for neat and doped flames, respectively. The contours of the model and experimental NO profiles are similar. However, the modeled neat and doped profiles are shifted 2mm toward and 1mm away from the burner surface, respectively, as compared to the measured profiles. Moreover, the modeled NO profile for the doped flame peaks at 9 mm and gradually decreases thereafter, in contrast to the experimental profile which plateaus at 9-10 mm. This discrepancy is related to an over prediction of NH₃ in the post flame region by the model. Results of equilibrium calculations for NO at 13.75 mm are three orders of magnitude lower than both flame code and experimental values, indicating that NO in the post flame region is well above its equilibrium concentration.

Presented in Figure 2A is a reaction pathway diagram depicting the nitrogen chemistry occurring in the neat H₂/N₂O/Ar flame. The numbers in parentheses are the relative integrated rates (0-13.75 mm) of the various reactions normalized to 100 for the reaction $\text{NH}_3 + \text{OH} = \text{NH}_2 + \text{H}_2\text{O}$ in the doped flame. As seen from Figure 2A, the reaction $\text{N}_2\text{O} + \text{H} = \text{N}_2 + \text{OH}$ represents the fastest step consuming N₂O to generate N₂. The reaction, $\text{N}_2\text{O} + \text{M} = \text{N}_2 + \text{O} + \text{M}$, is also important. NO is formed predominantly from the reaction $\text{N}_2\text{O} + \text{H} = \text{NO} + \text{NH}$. The formation of NO from NH, via the HNO intermediate, and reaction $\text{N} + \text{OH} = \text{NO} + \text{H}$ are also significant.

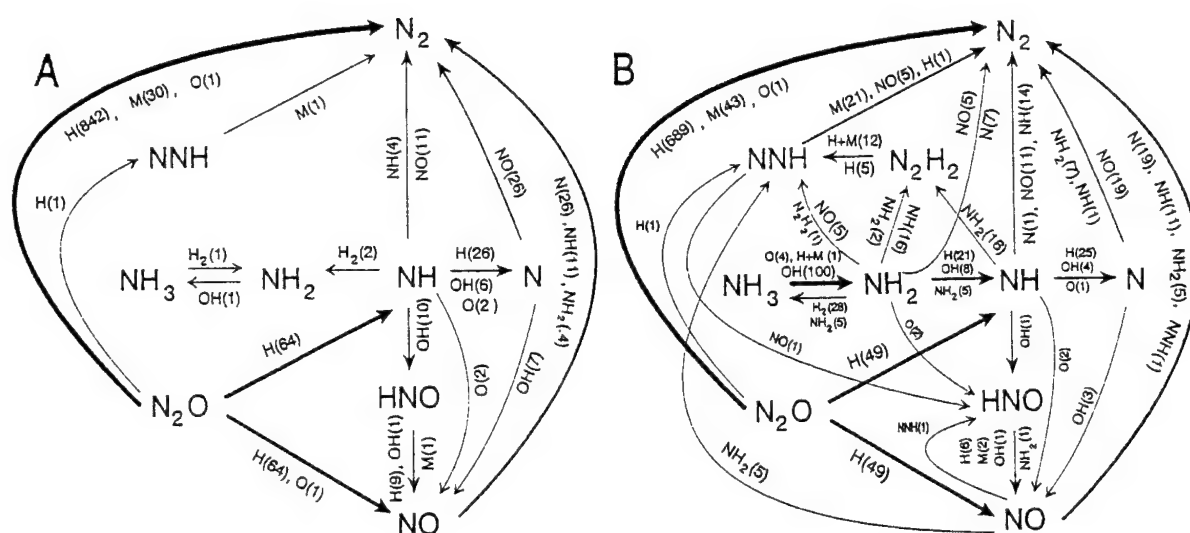


Figure 2. Pathway diagrams of neat (A) and doped (B) $\text{H}_2/\text{N}_2\text{O}$ flame at 13.75 mm.

The partial conversion of NO to N_2 occurs predominantly by reactions involving N and NH . Not surprisingly, the species NH_3 , NH_2 , NNH , and N_2H_2 do not play major roles in the neat flame.

Figure 2B shows the reaction pathway diagram for the NH_3 -doped flame. The addition of 4% NH_3 increases the number of reactions and species having major effects on the flame. As expected, the addition of the dopant causes only a small perturbation in the reactions involving the major species as compared to those in the neat flame. However, the addition of NH_3 does significantly reduce the NO concentration. This results primarily from two effects: (1) a slight decrease in NO formation from reactions $\text{N}_2\text{O} + \text{H} = \text{NO} + \text{NH}$ and $\text{N} + \text{OH} = \text{NO} + \text{H}$, and (2) an increase in NO consumption by the NH_2 radical. NH_2 can convert NO to N_2 directly, or indirectly via the NNH intermediate. It should be specifically noted that the lowering of NO concentration in the doped flame is primarily chemical in origin, rather than due to the slight change in temperature profiles. This was proven by switching the input temperature profiles in the calculations. Switching the profiles causes only a slight change in the predicted NO profile, particularly in the burnt gas concentrations for either the neat or doped flame.

The sensitivity coefficients for NO to various reaction rate coefficients were calculated for both neat and NH_3 -doped $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ flames at 13.75 mm. For both flames, the NO concentration has a strong negative sensitivity to $\text{N}_2\text{O} + \text{H} = \text{OH} + \text{N}_2$ and strong positive sensitivity to the reaction $\text{N}_2\text{O} + \text{H} = \text{NH} + \text{NO}$. This result is clearly due to the competition between the two channels for the $\text{N}_2\text{O} + \text{H}$ reaction, one forming NO , the other N_2 . For the neat flame, NO exhibits lesser but still important sensitivities to reactions $\text{NH} + \text{OH} = \text{H}_2\text{O} + \text{N}$, $\text{H}_2 + \text{OH} = \text{H}_2\text{O} + \text{H}$, $\text{NO} + \text{N} = \text{N}_2 + \text{O}$, $\text{NO} + \text{H} = \text{N} + \text{OH}$, $\text{N}_2\text{O} + \text{M} = \text{N}_2 + \text{O} + \text{M}$, and $\text{NH} + \text{NO} = \text{N}_2 + \text{OH}$. For most of these reactions the trends are similar for the doped flame. However, one notes that in comparison, the doped flame shows much stronger sensitivities than the neat flame to several additional reactions, particularly those reactions involving NH_2 , NH , or N . This is not surprising since the pathway diagrams indicate that these species play key roles in the consumption of NO . The sign of some of the sensitivity coefficients may be understood in terms of the competition between $\text{NH}_x + \text{NO}$ reactions, which result in the destruction of NO , and $\text{NH}_x + \text{NH}_y$ reactions, which ultimately result in conversion of the NH_3 to N_2 without removal of NO . For example, reactions $\text{NH}_2 + \text{NO} = \text{NNH} + \text{OH}$ and $\text{NH}_2 + \text{NO} = \text{N}_2 + \text{H}_2\text{O}$ have negative coefficients because they ultimately convert NO to N_2 . In contrast the $\text{NH}_2 + \text{NH} = \text{N}_2\text{H}_2 + \text{H}$ and $\text{NH} + \text{NH} = \text{N}_2 + 2\text{H}$ reactions merely convert two NH_x molecules ultimately to N_2 without destroying NO , thus exhibiting a positive NO sensitivity coefficient. It appears that this competition for NH_x molecules is the reason for the leveling off of the effectiveness of NH_3 in reducing the NO which is observed experimentally. As the concentration of NH_3 is

increased, the importance of $\text{NH}_x + \text{NH}_y$ reactions increases.

Conclusion

A combined experimental and modeling flame structure study of neat and NH_3 -doped $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ flames has been performed. Species mole fractions and temperature profiles were recorded using MB/MS and thermocouple techniques, respectively. The modeling studies consisted of both equilibrium and flame code calculations. A chemical mechanism for the system was developed from a critical literature review and employed for modeling the species profiles. The modeled profiles of the majority species agree very well with the experimental profiles for the neat flame and reasonably well for the doped flame. Quantitatively, the agreement between the modeled and experimental NO profiles is good for both the neat and doped flames. The modeled profile shows a 55% reduction in the NO mole fraction in the post flame region when 4% NH_3 is added to the neat flame compared to a 45% reduction in the measured profile. The modeled and experimental NO profiles for the neat flame and the experimental profile in the doped flame exhibit plateaus in the post flame region. However, the modeled NO profile in the doped flame exhibits a post flame decay. In addition, the model overpredicts the NH_3 mole fraction in the post flame region. Rate analysis shows conclusively that this overprediction is the primary cause of the predicted NO decay. Calculations indicate that these discrepancies are due to the model and not the experiment, suggesting that refinements in the chemical mechanism are necessary. Rate and sensitivity analysis reveal that the dopant decreases slightly the amount of NO formed by the reactions $\text{N}_2\text{O} + \text{H} = \text{NO} + \text{NH}$ and $\text{N} + \text{OH} = \text{NO} + \text{H}$, and increases its consumption via reaction with NH_2 , which ultimately leads to conversion of NO to N_2 . The efficacy of NH_3 in reducing NO is stymied at high NH_3 concentrations, an effect likely due to the increased role of $\text{NH}_x + \text{NH}_y$ reactions which convert NH_3 to N_2 without removing NO.

Acknowledgments

This work was supported by the ARL Combustion Research mission program and in part by a grant of HPC (High Performance Computing) time from the DoD HPC Shared Resource Center, Corps of Engineers Waterways Engineering Station, Cray Y-MP8.

References

1. Sausa, R., Anderson, W., Dayton, D., Faust, C., and Howard, S., "Detailed Structure Study of a Low Pressure, Stoichiometric $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ Flame," *Combust. and Flame*, Vol 94, pp. 407-425, 1993;
2. Dayton, D., Faust, C., Anderson, W., and Sausa, R., "Flame Structure Study of a Lean $\text{H}_2/\text{N}_2\text{O}/\text{Ar}$ Flame Employing MB/MS and Modeling", *Combust. and Flame*, Vol 99, pp. 323-330, 1994.
3. Miller, J., and Bowman, C., "Mechanism and Modeling of Nitrogen Chemistry in Combustion," *Prog. Enrg. Combust. Sci.*, Vol 15, pp. 287-338, 1989.
4. Allen, M., Yetter, R., and Dryer, F., "The Decomposition of Nitrous Oxide at $1.5 \leq P \leq 10.5$ Atm and $1103 \leq T \leq 1173$ K," *Int. J. Chem. Kinet.*, Vol 27, pp. 883-909, 1995.
5. Glarborg, P., Johnson, J., and Dam-Johnsen, K., "Kinetics of Homogeneous Nitrous Oxide Decomposition," *Combust. Flame*, Vol 99, pp. 523-532, 1994.
6. Mertens, J., Chang, A., Hanson, R., and Bowman, C., "A Shock Tube Study of the Reactions of NH with NO , O_2 , and O ," *Int. J. Chem. Kinet.*, Vol 23, pp. 173-196, 1991, and references therein.

Engineering and Environmental Sciences

Laser Velocimetry and Doppler Global Velocimetry Measurements of Velocity Near the Empennage of a Small-Scale Helicopter Model

Susan Althoff Gorton*
JRPO-AFDD, ATCOM

James F. Meyers
National Aeronautics
and Space Administration

John D. Berry
JRPO-AFDD, ATCOM

Langley Research Center
Hampton, VA 23681

Abstract

A test program was conducted in the NASA Langley 14- by 22-Foot Subsonic Tunnel to measure the flow near the empennage of a small-scale powered helicopter model with an operating tail fan configuration. Velocity profiles were measured with three-component Laser Velocimetry (LV) one chord forward of the horizontal tail for four forward speeds to evaluate the effect of the rotor wake impingement on the horizontal tail angle of attack. These velocity data indicate the horizontal tail can experience unsteady downwash angle variations of over 30° due to the rotor wake influence. The horizontal tail is most affected by the rotor wake above speeds of 23 m/s (44 knots). Three-component velocity measurements of the flow on the inlet side of the fan were made for a low-speed flight condition using both LV and a promising, non-intrusive, global, three-component velocity measurement technique called Doppler Global Velocimetry (DGV). The velocity data show an accelerated and non-uniform flow into the fan. DGV shows promise as an evolving tool for rotor flowfield diagnostics.

1. Introduction

As rotor and fuselage designs become more integrated, compact, and complex, close rotor wake-fuselage interactions and interference play an increasingly important part in the performance characteristics of rotorcraft. Sheridan and Smith (1979) attribute the importance of interactional effects for modern helicopters to increased disk loading, more compact designs, low level flight requirements, and the increased requirement for directional trim after the loss of the tail rotor which results in larger vertical tail surfaces. These effects are especially important in the design and placement of the anti-torque system and the horizontal and vertical stabilizers as documented in Prouty and Amer (1982).

Much work has already been done experimentally and analytically to define the interaction effects between the rotor and the fuselage. More limited is the amount of experimental data available for analyzing the main rotor/anti-torque interactions. As advanced configurations such as the RAH-66 are designed and manufactured with sophisticated anti-torque devices, there is a need for high-quality experimental data to support the development of analytical models which have the flexibility to model these types of configurations. Torok and Ream (1993) specifically cite the difficulty in predicting unsteady empennage loads at speeds below 22 m/s (40 knots). While Moedersheim and Leishman (1995) provide experimental pressure data at model scale for a generic T-tail empennage, and Keys et al. (1991) discuss the tremendous amount of testing involved in the Light Helicopter (LH) design process, there does not appear to be specific information in the literature on the velocities in the flowfield near an operating tail fan.

In order to investigate the rotor wake/fuselage/empennage interactions near the empennage of a powered small-scale helicopter with an operating tail fan and a T-tail, the U. S. Army Joint Research Program Office, Aeroflightdynamics Directorate, in cooperation with the NASA Langley Measurement Sciences and Technology Branch, recently conducted a wind tunnel test program in the 14- by 22-Foot Subsonic Tunnel.

2. Model and Instrumentation

The test program was conducted in the Langley 14- by 22-Foot Subsonic Tunnel using the Army's 2-Meter Rotor Test System (2MRTS), a four-bladed, 15-percent scale rotor, a fuselage model representative of the RAH-66, and the tunnel's three-component Laser Velocimetry (LV) system. In addition, a new optical flow measurement technique, Doppler Global Velocimetry (DGV), was applied for the first time to a rotorcraft flowfield during this test program.

The 14- by 22-Foot Subsonic Tunnel is a closed-circuit, atmospheric wind tunnel designed for the low-speed testing of powered and high-lift configurations. This investigation was conducted with the tunnel in the open test section configuration to allow complete optical access to the rotor flowfield. Figure 1 shows the 2MRTS ready for testing in the tunnel. The LV system in operation is also visible in the photograph. The rotor system which was installed on the 2MRTS was a 4-bladed, articulated hub with blades that closely matched the airfoils, planform, and twist of the RAH-66 blades. The anti-torque device of the configuration was modeled by a tip-driven, 20.3-cm (8-in) diameter, 22-bladed fan mounted in the tail fan duct. The fan configuration is shown in Figure 2. As the photograph shows, the fan duct section was painted black to minimize the optical reflections from the surface.

The LV system was a three-component system operating in the backscatter mode. The downstream and vertical components of velocity are measured by the optics which are located on the side of the tunnel, out of the flow; the lateral crossflow component of velocity is measured by the optics which are located beneath the tunnel floor. As can be seen in Figure 1, the third component beams originating beneath the flow were angled at 33° to the vertical to access the inflow area of the canted tail fan.

Except for its long focal length and zoom lens assembly, the system operated as a standard fringe-based LV system; polystyrene particles ($1.7\text{ }\mu\text{m}$) suspended in an alcohol and water mixture were used to seed the flow. The velocity data were acquired using Frequency Domain Processors (FDP's) to maximize the signal to noise ratio of the measurement signal. The LV data acquisition system was designed to acquire rotor azimuth position in addition to the velocity measurements so that an "azimuthal history" of the velocity could be reconstructed in the post-processing of the velocity data.

DGV is a fairly new technique to measure three-components of velocity which has been mainly applied to fixed-wing studies. To date, this technique has been used to acquire only steady-state data; however, the extension of the technique to capture the unsteady rotor flowfield is underway at Langley. This effort is a jointly funded project between the Army and NASA. As a first step in applying DGV to rotorcraft, the technique was used during this test program to measure three-component mean (or steady-state) velocity near the fan tail. This established the feasibility of using DGV in the facility for the types of low-speed and reversed flows which occur in rotor wakes.

The DGV system which was used for this test program was based on the theories described in Meyers (1994). For this test program, three sets of two cameras each were used; each camera pair measured a component of velocity. Although the camera pairs did not measure orthogonal velocity components, post-test processing transformed the velocity measurements into the standard u , v and w components.

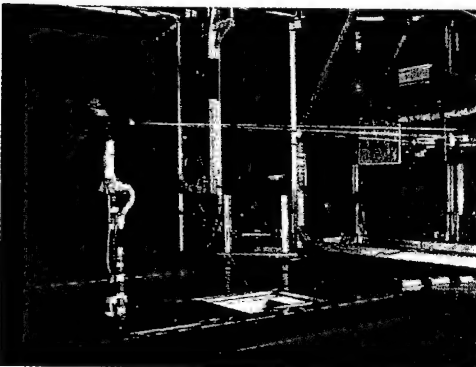


Figure 1. Model and LV installed in tunnel.

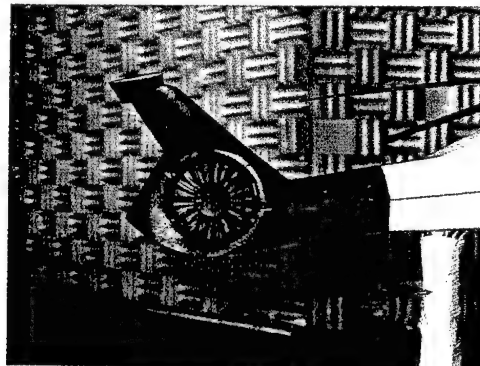


Figure 2. Tail fan configuration.

A light sheet produced by a single frequency Argon-ion laser was projected into the flow. This sheet defined the measurement plane for the three-component DGV velocity measurements. The DGV measurements were made in the same plane as the LV measurements so that detailed comparisons of the two techniques could be made. Propylene glycol smoke was injected into the flow from the settling chamber of the tunnel. As the smoke passed through the sheet, it scattered light that was Doppler-shifted in optical frequency proportional to the particle velocity.

At each camera pair location, an iodine vapor cell was placed in front of the signal video camera to attenuate the collected scattered light in proportion to the shift in the laser light optical frequency. A second camera, without an iodine cell, was used to provide a reference image of the scattered light intensity distribution emitted by the smoke passing through the light sheet. Normalization of the signal image by this reference image removed spatial variations in light intensity. The resulting image amplitude distribution was a map of the velocity flowfield in the locations illuminated by the laser light sheet.

3. Discussion of Results

Velocities were measured in two regions with both the main rotor and the tail fan operating. In the first region, forward of the horizontal tail, velocity was measured with the LV system for four forward speeds. In the second region, on the inlet side of the tail fan, velocity was measured with both LV and DGV.

3.1 Horizontal Tail

The average downwash angle at the tail is presented in Figure 3. As expected, the downwash decreases with increasing speed as the rotor wake is swept downstream. In Figure 4, the unsteady downwash angle is shown for four speeds. These data show that the tail can experience angle variations of over 30° occurring at frequencies of 4/rev. By analyzing the frequency content of the velocity time history, it is possible to determine the position of the rotor wake relative to the tail for each speed. Figure 5 shows the results of this analysis. The rotor wake impinges on the horizontal tail at speeds above 23.5 m/s (44 knots).

3.2 Tail Fan

Figure 6 presents contour plots in the fan system coordinates of the average downstream velocity, u , the lateral (perpendicular to the fan) velocity, v , and the vertical (parallel to the fan) velocity, w . These are presented for a forward speed of 18.2 m/s (33 knots) and an equivalent full-scale vehicle weight of 33,362N (7500 lbs). Note the accelerated flow at the forward section of the duct. The data indicate the flow is separated along the upstream lip of the tail fan duct.

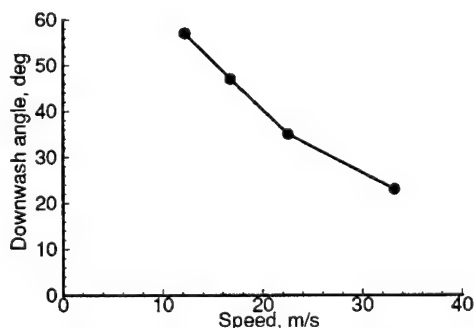


Figure 3. Average downwash angle.

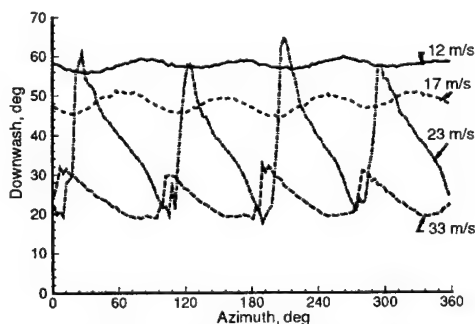


Figure 4. Unsteady downwash angle, one chord forward, one chord to the right, and 0.13 chord below horizontal tail centerline.

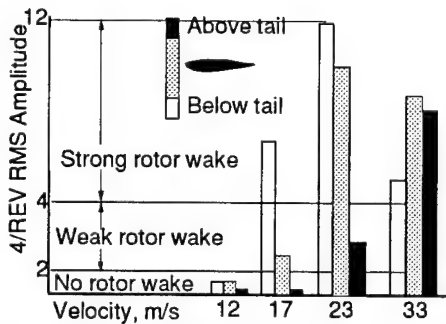


Figure 5. Wake impingement on horizontal tail.

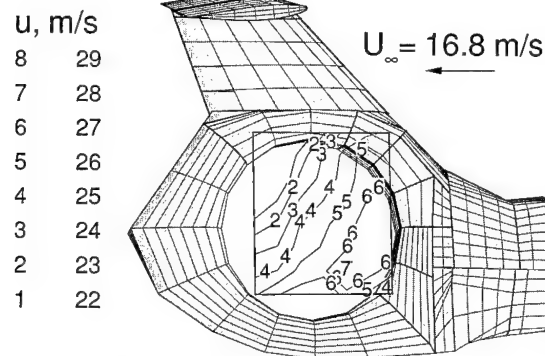
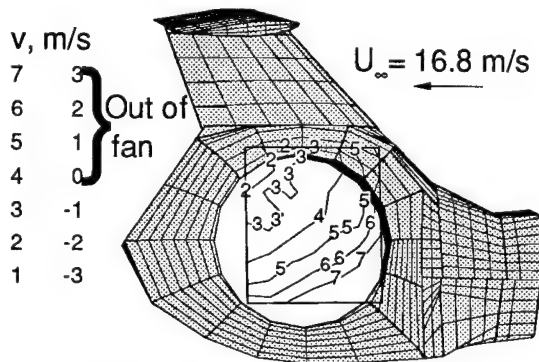
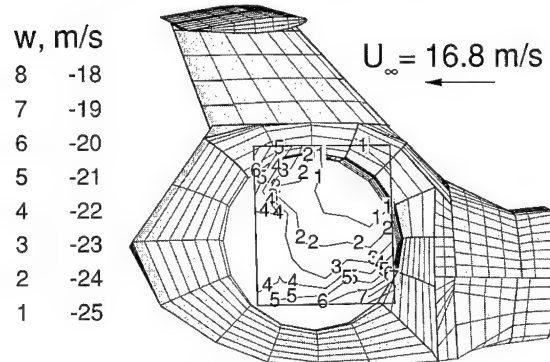
Figure 6a. Streamwise velocity, u .Figure 6b. Normal (inflow) velocity, v .Figure 6c. Vertical velocity, w .

Figure 6. Contour plots of LV average velocity data near the tail fan.

In Figures 7-8, the DGV data are presented. The data in Figure 7 are for the individual components of velocity in the wind tunnel coordinate system; the dark areas on the right of the figures are the absence of smoke in the image. These plots illustrate how DGV maps the entire velocity field illuminated by the smoke in the light sheet. As these velocity maps are difficult to assess, especially in gray scale, Figure 8 presents the DGV data for a single slice horizontally through the measurements for each velocity component. The location for this slice of a single row of pixels is shown on Figure 7 by arrows on each of the velocity maps. The LV data for the same configuration are also presented on the plots in Figure 8. Note that the DGV data provides a tremendous increase in the resolution of the velocity measurement locations; there are over 400 DGV measurements to compare to 7 LV measurements in the same horizontal line.

Although the DGV data do not match exactly with the LV data, there is enough similarity to encourage the continued development of the global velocity technique. The DGV technique requires some additional refinement and improvements before it is established as a reliable, accurate tool for rotorcraft; however, the immense potential payoff of increased efficiency in flowfield measurement capability is worth the investment.

Figure 7a. u -component.Figure 7b. v -component.
Figure 7. DGV velocity maps.Figure 7c. w -component.

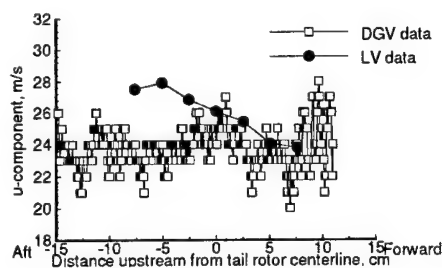
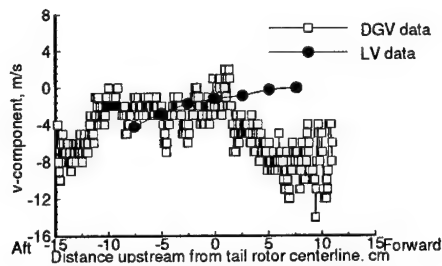
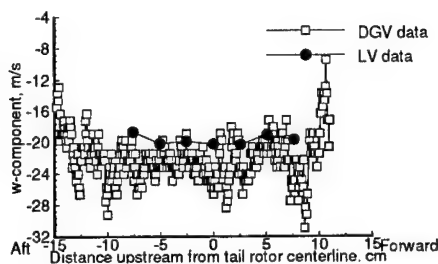
Figure 8a. Streamwise velocity, u .Figure 8b. Normal (inflow) velocity, v .Figure 8c. Vertical velocity, w .

Figure 8. Comparison of DGV and LV data.

4. Conclusions

In order to investigate the rotor wake/fuselage/empennage interactions near the empennage of a powered small-scale helicopter with an operating tail fan and a T-tail, the U. S. Army Joint Research Program Office, Aeroflightdynamics Directorate, in cooperation with the NASA Langley Measurement Sciences and Technology Branch, recently conducted a wind tunnel test program in the 14- by 22-Foot Subsonic Tunnel. Velocity data were acquired forward of the horizontal tail for four flight conditions, documenting the unsteady downwash near the horizontal tail. Velocity data were also obtained on the inlet side of the fan for one flight condition, providing information about the inflow into the tail fan. The major conclusions from this study are: 1) There is an accelerated flow pattern near the fan tail. 2) The horizontal tail surface experiences large changes (over 30°) in the unsteady sidewash and downwash angles due to the influence of the rotor wake. The horizontal tail is most affected by the rotor wake above speeds of 23.7 m/s (44 knots). 3) A new flow measurement technique, Doppler Global Velocimetry, shows promise for improving non-intrusive global velocity measurement productivity.

References

1. Sheridan, P. F., and Smith, R. P., *Interactional Aerodynamics--A New Challenge to Helicopter Technology*, American Helicopter Society 35th Annual Forum, Washington, DC, May 1979.
2. Prouty, R. W., and Amer, K. B., *The YAH-64 Empennage and Tail Rotor--A Technical History*, American Helicopter Society 38th Annual Forum, Anaheim, CA, May 1982.
3. Torok, M. S., and Ream, D. T., *Investigation of Empennage Airloads Induced by a Helicopter Main Rotor Wake*, American Helicopter Society 49th Annual Forum, St. Louis, May 1993.
4. Moedersheim, E., and Leishman, J. G., *Investigation of Aerodynamic Interactions Between a Rotor and a T-Tail Empennage*, Presented at the American Helicopter Society Aeromechanics Specialist Meeting, Fairfield County, CT, October 1995.
5. Keys, C., Sheffler, M., Weiner, S., and Heminway, R., *LH Wind Tunnel Testing: Key to Advanced Aerodynamic Design*, American Helicopter Society 47th Annual Forum, Phoenix, May 1991.
34. Meyers, J. F., *Development Of Doppler Global Velocimetry For Wind Tunnel Testing*, AIAA 18th Aerospace Ground Testing Conference, Paper 94-191490, Colorado Springs, CO, 1994.

Modeling of the Mixing/Combustion Process in a Quiescent Chamber D.I. Diesel Engine

P. Schihl*, W. Bryzik, Schwarz, E., and E. Danielson
U.S. Army Tank-Automotive RDE Center (TARDEC)
Warren, MI., 48397-5000

Abstract

A global mixing and combustion model has been developed for a low swirl, direct injection (DI) diesel engine. It includes relevant injection parameters, the engine operating condition (speed and load), and in-cylinder thermodynamic state through a variety of submodels which encompass *only one combustion chamber dependent constant*. Other similar models possess much more empiricism and additional constants. As a corollary, a spray penetration model was developed and validated by data from three independent sources - Sandia National Labs and the Universities of Wisconsin and Hiroshima. Similarly, the combustion model has also been compared with experimental data generated at TARDEC (under four engine operating conditions) and shows a great deal of promise. Future plans are being made to further validate the model for another quiescent type diesel engine in conjunction with continued development of the injection rate and wall heat loss modules.

1. Introduction

The overall combustion process in a DI diesel engine is comprised of a number of key individual phenomena. Of utmost importance are the fuel spray penetration, subsequent mixing, and ultimately fuel burning subprocesses. Models have been developed for each of these key components and are briefly described with more attention paid particularly to the mixing and fuel burning rate models.

1.1 Spray Penetration

As outlined by Schihl, Bryzik, and Atreya [1996a], the injected fuel jet is modeled as a cone which penetrates the surrounding medium with some average cone angle and injection velocity; see figure 1. The conservation of linear momentum is applied axially to the cone with the following assumptions:

- (1) minimal momentum flux induced axially on the cone by entrainment
- (2) minimal viscous forces on cone periphery
- (3) negligible body force
- (4) well-mixed tip, i.e. $x/d \gg 1$, $\rho_{tip} \approx \rho_m$
- (5) uniform injection velocity profile
- (6) polynomial velocity profile at cone tip
- (7) constant discharge coefficient, c_d

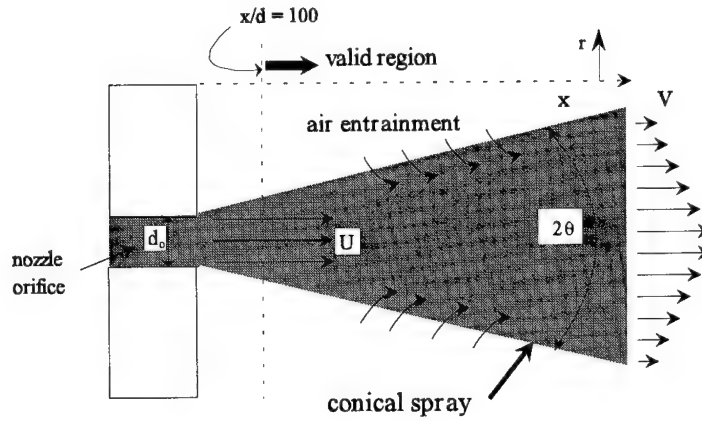


Figure 1 : The cone model

where ρ_∞ is the average mixing chamber density. (The first two assumptions are scaling arguments; the momentum influx associated with entrainment and the viscous force on the cone are assumed much smaller than the injected momentum and propagating tip momentum.) After evaluating the momentum equation integral, the penetration distance is given by:

$$x(t) = \left[\frac{2c_d}{\tan \theta} \right]^{1/2} \left(\frac{\Delta P}{\rho_\infty} \right)^{1/4} (d_o t)^{1/2} \quad (1)$$

where θ is the average cone angle, ΔP is the pressure drop across the injector, d_o is the orifice diameter, and t is the time after injection. A modified Ranz as proposed by Ruiz and Chigier [1991] is employed for calculating the cone angle:

$$\tan \theta = \frac{4\pi}{A} \left(\frac{\rho_\infty}{\rho_1} \right)^{1/2} f \left[\frac{\rho_1}{\rho_\infty} \left(\frac{Re_1}{We_1} \right)^2 \right] \left(\frac{Re_1}{We_1} \right)^{-1/4} \quad (2)$$

where A is a constant associated with the nozzle and ρ_1 is the injected fuel density; the function f is given by Reitz [1979].

1.2 Fuel-Air Preparation

The mixing rate in a quiescent chamber, DI diesel engine is controlled by various processes. The intake flow during the induction stroke and part of the compression stroke determines the initial state of the flow field within the cylinder after intake valve closing (IVC). From IVC to exhaust valve opening (EVO), the in-cylinder flow field is dependent on the rate of compression and expansion (engine speed and geometry), bowl re-entry flow (squish), cylinder wall heat loss or gain, mixture compressibility, fuel injection rate, combustion, and others.

Rate of compression or expansion affects turbulent length and time scales by distorting the in-cylinder flow field while squish and fuel injection rate modulate the flow field by momentum transfer. Wall heat loss, mixture compressibility, and combustion can cause substantial changes to various physical properties and turbulence scales of the flow field as well. These various processes are highly complex and occur in a three dimensional manner.

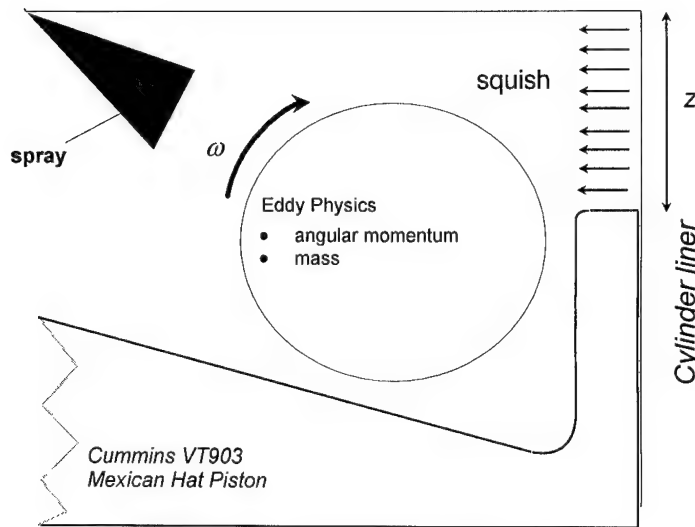


Figure 2 : Large-scale mixing model

In order to model this problem in a physically interpretive but relatively simple manner, the highly complex three-dimensional process must be simplified to a more workable format. First of all, in typical diesel-type injectors, there are multiple fuel sprays propagating into each engine cylinder. Imagine dividing the combustion chamber into a number of planes, equal to the number of nozzle orifices, which each contain one penetrating fuel spray. Next, assume that each spray plume can be modeled independently; at this point, the three-dimensional problem has been reduced to a two-dimensional problem with each injected spray plume isolated from the other spray plumes.

Based on this simple argument, the proposed mixing model is shown schematically in figure 2 and is briefly described as follows. After IVC, the in-cylinder flow field is composed predominately of one large-scale eddy which has some initial mass and angular momentum. The existence of smaller, recirculating flows is ignored since it is believed the large-scale eddy is primarily responsible for mixing of the fuel and air. As the piston moves toward top dead center (TDC), the scale of the eddy decreases and the rate of re-entry flow (squish) increases - the squish adds mass to the eddy causing a reduction in eddy turnover rate. Near TDC, fuel is injected into the combustion chamber causing an increase in eddy angular momentum. After TDC, the cylinder expands, resulting in a corresponding increase in eddy size and reversal of squish flow until EVO - the net effect of expansion is to decrease the eddy turnover rate. Of course, throughout the whole process from IVC to EVO, the eddy strength is decaying due to viscous dissipation.

1.2.1 Formulation

Analysis of the mixing rate is based on a very simple philosophy. Basically, the large-scale eddy angular momentum and mass are conserved with time as shown qualitatively below:

$$\frac{d}{dt} (\text{eddy angular momentum}) = -\text{dissipation} + \text{production} \quad (3)$$

$$\frac{d}{dt} (\text{eddy mass}) = \text{net entrainment} + \text{injection} \quad (4)$$

The production of angular momentum is predominately due to squish flow and fuel injection and initially the eddy is assumed to contain all the mixing chamber gaseous mass excluding the squish region such that its physical content is only modulated by injection and squish. The exact details of the formulation are given by Schihl, Bryzik, and Atreya [1996b]. Essentially, the turnover or mixing rate is given by a nonlinear differential equation which includes various physical phenomena:

$$\frac{d\omega}{dt} = -\frac{\omega}{m_e} \frac{dm_e}{dt} - 2\frac{\omega}{R} \frac{dR}{dt} - a\omega^2 \exp(-a\omega \delta t) + 2\left(\frac{\rho_{sq} A_{sq} l_{sq}}{m_e}\right) \left(\frac{U_{sq}}{R}\right)^2 + \frac{\dot{P}_{inj}}{\frac{1}{2}m_e R^2} \quad (5)$$

where m_e and R are the instantaneous eddy mass and radius, a is the geometrically dependent dissipation constant, δt is the computational time step, A_{sq} and l_{sq} are the squish area and length scale, and \dot{P}_{inj} is the angular momentum production due to injection. The first term on the right-hand side represents induction/abduction of fuel and air into or out of the eddy while the second term allows for length scale modulation. The last two entities include the effects of both squish and fuel injection.

1.3 Combustion

Before autoignition, the injected fuel jet penetrates into the combustion chamber inducing momentum onto the in-cylinder flow field. Throughout this process, a shear layer is formed on the periphery of the jet, consisting of varying proportions of fuel-air mixture. Eventually, the local thermodynamic conditions reach a state allowing for autoignition in either one or multiple locations. If regions around 'enough' ignition locale are conducive to establishment of a flame, the combustion process proceeds in a typical fashion. In this case, the flame spreads very quickly throughout the ignitable shear layer consuming the fuel in a rapid manner while, in the meantime, the rest of the injected fuel not encompassed by the shear layer, continues to mix with the surrounding oxidizer as dictated by the flow field. Inevitably, the fuel in the shear layer is consumed and the fuel-oxidizer mixing rate becomes solely responsible for the rate of fuel consumption. This description is the basis for the proposed two-phase combustion model.

1.3.1 Premixed Phase

The injected fuel jet propagates into the gaseous medium forming a distorted conical shape. A combustible mixing layer is formed on the spray periphery which initially increases monotonically along the plume axis but eventually decreases to a fairly thin layer at the plume tip. The predominately air mixture within the combustion chamber is compressed at a rapid rate and correspondingly, the in-cylinder temperature increases substantially from its initial value after IVC to a substantially higher level just before the onset of injection (which typically occurs near TDC for most diesel engines). The resulting hot air is mixed with the injected fuel forming the aforementioned shear layer around the spray perimeter. A thicker region corresponds locally to larger amounts of hot air mass but smaller radial concentration gradients. Eventually, the temperature and species concentrations are conducive to

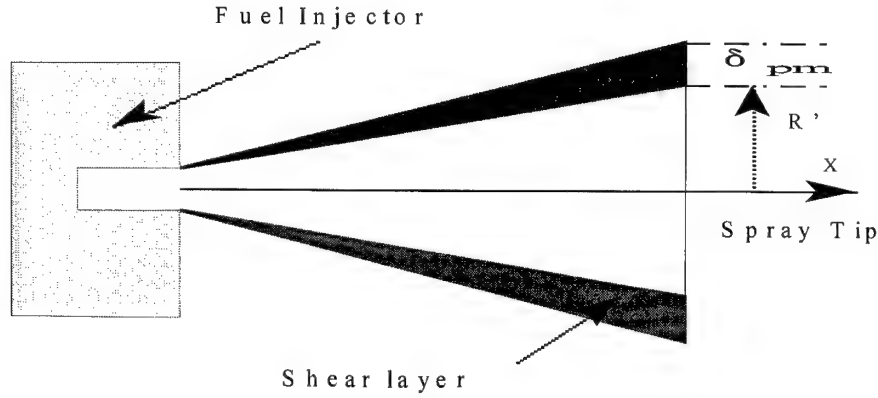


Figure 3 : Premixed phase combustion model

autoignition at either one or more sites. Arguably, this phenomena might occur near the maximum shear layer thickness due to higher temperature and lower strain rates. Thinner regions of the mixing layer have larger thermal and concentration gradients. However, they too will subsequently approach autoignition conditions. This physical picture does not exclude multiple ignition sites as observed by Edwards, Siebers and Hoskin [1992] under various combustion chamber initial conditions. However, their occurrence is of little relevance due to their short time-scale compared with the ignition delay and ensuing flame spread rate.

Essentially, this aforescribed process is modeled as the burning of a shear layer on a propagating conical jet (see figure 3). Ignition is assumed to take place at the tip of the cone with the resulting flame propagating back toward the injector. The entrained, unburned fuel-air mixture is assumed to burn over some characteristic time scale as dictated by the in-cylinder turbulence level and laminar flame velocity. After exhaustion of the fuel in the shear layer the premixed phase reaches termination.

The prescribed fuel burning rate model is very similar to that developed by Tabaczynski [1977] for spark ignition engines but adapted for the fuel injection process:

$$\frac{dm_{pb}}{dt} = \frac{m_e - m_{pb}}{\tau} \quad (6)$$

where m_{pb} is the fuel burned, m_e is the entrained fuel-air mixture, and τ is the characteristic burning time given by:

$$\tau = \frac{\delta_t}{S_l} \quad (7)$$

$$\frac{\delta_t}{l} \propto Re_l^{-1/2} \quad (8)$$

where Re_l is the turbulent Reynolds number, δ_t is the Taylor scale, and S_l is the laminar flame speed. The rate of fuel-air entrained in the flame front, the turbulence level, and flame front area are given by:

$$\frac{dm_{en}}{dt} = \rho_u A_f (u' + S_l + U_{jet}) \quad (9)$$

$$u' = l \omega \quad (10)$$

$$A_f = \pi [(R' + \delta_{pm})^2 - R'^2] \quad (11)$$

where l is the eddy length scale (two times the eddy radius) and U_{jet} is the jet velocity at the current flame front location.

1.3.2 Mixing Controlled Phase

The onset of the second phase of combustion is assumed to begin once the first parcel of fuel-air mixture not contained in the premixed phase reaches flammable limits after autoignition. The fuel burning rate during this regime is dominated by the mixing rate:

$$\frac{dm_{db}}{dt} = C_1 (m_a - m_{db}) \omega \quad (12)$$

where C_1 is some constant on the order of 1, m_a is the available fuel found via the mixing rate, and m_{db} the fuel consumed during only the mixing controlled phase.

The fuel consumption rate during the overall combustion process is represented by contributions from both the premixed and mixing controlled phases. Thus, it is given by:

$$\frac{dm_b}{dt} = \frac{dm_{pb}}{dt} + \frac{dm_{db}}{dt} \quad (13)$$

2. Solution for the Mixing and Fuel Burning Rates

A simple scheme is used to solve the governing equations for the eddy mixing rate, ω , and mass, m_e . Starting with the initial conditions, a first order Taylor series is employed to obtain the mixing rate and eddy mass at later times as given by equations 14 and 15:

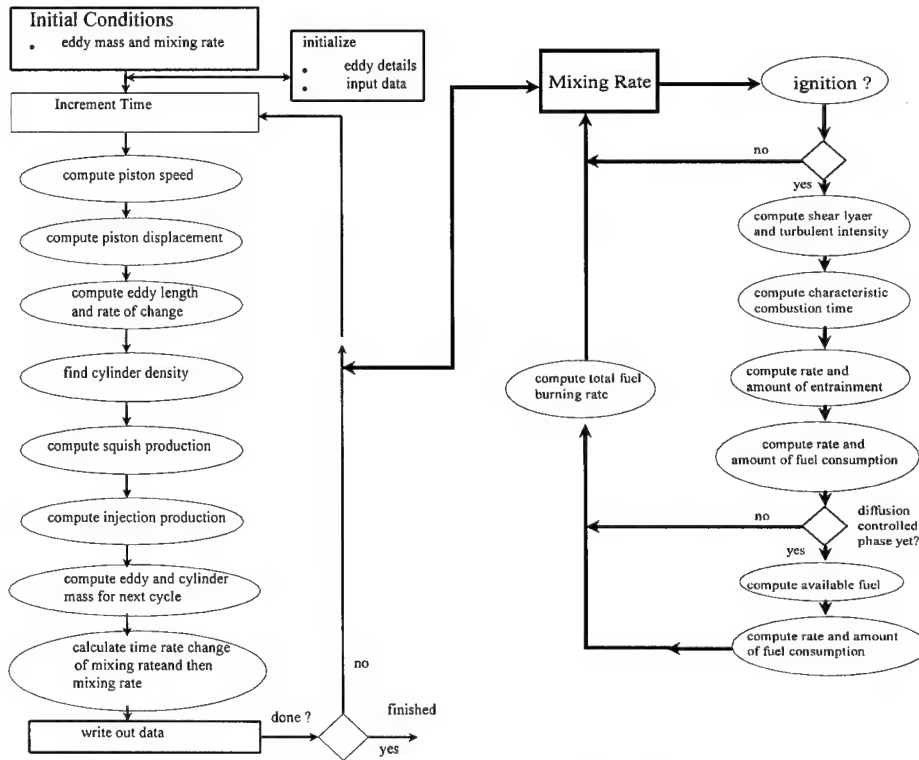


Figure 4 : Solution algorithm for determination of mixing and fuel consumption rates

$$m_e(i+1) = m_e(i) + \frac{dm_e(i)}{dt} \delta t \quad (14)$$

$$\omega(i+1) = \omega(i) + \frac{d\omega(i)}{dt} \delta t \quad (15)$$

where i represents some point in time. A very modular FORTRAN program was written to solve the governing mass and momentum equations. The solution algorithm is shown in figure 4.

First, the initial mixing rate and eddy mass must be estimated for the given engine operating condition. Then, the relevant geometric and physical parameters of the combustion chamber are initialized along with the initial crank angle position of the engine. After these preliminary issues are addressed, the piston speed and displacement along with the eddy length scale and rate of scale change are calculated from the given engine operating conditions. The average in-cylinder density is then computed from the current combustion chamber mass and volume in order to find the squish production rate. Next, the fuel injection production term is calculated and then the eddy and cylinder mass are updated for the next computational cycle. Finally, the mixing rate is determined

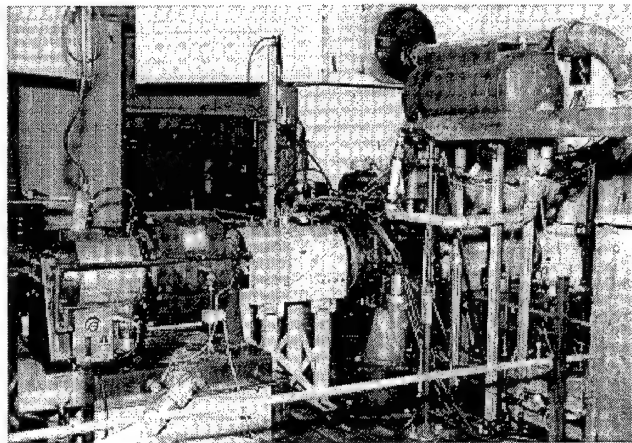


Figure 5 : TARDEC, in-house research engine setup

and any output is stored to a data file. If the simulation is unfinished, time is incremented and the mixing rate and eddy mass are calculated in the same procedure as described above.

Once the mixing rate is found for each time increment the fuel burning rate is readily attainable through integration of the premix and mixing controlled phase combustion models - see figure 4 for solution process.

After solving for the mixing rate at a particular time increment, the fuel-air mixture is monitored for autoignition. If conditions are not right for this phenomena, the mixing rate is found for the next time increment and the process is repeated once again. With the onset of autoignition, the shear layer thickness and turbulent intensity are found in conjunction with a characteristic burning time. Then, the amount of fuel entrained into the flame front, and subsequently burned, is computed using the premixed phase entrainment and combustion models. Next, the diffusion burn time delay is calculated and compared with the current time increment. If it is too early for the diffusion controlled combustion phase, the mixing rate is calculated for the next time step. Once the mixing controlled phase begins, the amount of available fuel is determined for each time step and the consumption rate is directly found from the corresponding combustion model. Lastly, the total fuel consumption rate is determined from consideration of both combustion phases. This whole process is repeated until the simulation is complete, as defined by the time at which the exhaust valves close.

3. Experimental Setup

A production type, Cummins VT903 diesel engine was modified for specific combustion research as outlined by Schwarz, Reid, Bryzik, and Danielson [1993]. Basically, the front six cylinders were removed from the original engine block leaving only the remaining two rear cylinders intact. Furthermore, a 'slipper' piston (SP) was employed in one of the two prevailing cylinders in order to explicitly concentrate on the combustion process in one cylinder. The SP is a specially designed piston which maintains near ambient pressure within the given combustion chamber and also provides the necessary required dynamic balancing inherent for the special V-2 engine configuration. The intake air, coolant, and oil temperatures were controlled externally using various heaters. Fueling rate was also regulated separately from the engine. A photograph of the test cell configuration is shown in figure 5.

Pressure data was acquired for the baseline and ceramic coated pistons under six different engine operating conditions. Three speeds - 1600, 2100, and 2600 RPMs - were investigated at half and full load with turbocharging simulated via control of the intake air temperature. The resulting pressure traces were further analyzed using the thermodynamic based, cycle simulation software package developed by Chen [1989] in order to study the effect of load, speed, and piston surface on the gross apparent heat release rate profile. For the present study, though, only the baseline configuration is applicable.

4. Initial and Boundary Conditions

A number of key parameters associated with the proposed combustion model were estimated from the experimental data. These items included start of injection (SOI), ignition delay, end of injection (EOI), initial in-cylinder turbulence level, and the eddy dissipation constant. SOI was determined to occur approximately just before a ripple (due to injection) was observed in the combustion chamber pressure gradient as the piston approached TDC. For the two lower engine speeds it was difficult to determine SOI using this technique. Instead, for these conditions (1600 and 2100 RPMs), SOI was estimated by the amount of fuel contained in the injector cup. The EOI was assumed to occur just after TDC for all engine speeds as dictated by the cam lift. Ignition delay was defined as the time elapsed between SOI and the beginning of fuel burning as governed by the apparent gross heat release profile while initial turbulence level was determined from the intake air flow rate to the engine. In the latter case, the inducted air was assumed to pass through the two intake valves and subsequently evolve into a number of large scale eddies as dictated by the combustion chamber geometry. Finally, the eddy dissipation constant was determined by comparing the simulated fuel burning rates with the experimentally generated gross heat release profiles.

5. Results

For brevity and also demonstrating capability and trends, only the lower two engine speeds will be discussed even though many of the general observations are further validated for the higher engine speed condition.

The calculated eddy turnover or mixing rate is shown in figure 6. Generally, the initial turbulence level decays until squish and combustion chamber compression rates become more dominant on in-cylinder mixing. At this point the turnover rate begins to increase more rapidly (around -30°) but eventually these two phenomena become less important near TDC and the mixing rate begins to decrease. This is evident for the half load cases. For the full load conditions, injection occurs before this behavior is observable. After SOI, there is a rapid increase in the turnover rate until just before EOI. For the full load cases, mixing reaches a maximum level near 12° after TDC, while under half load conditions, a peak rate is attained closer to EOI. Initially, the injection process results in a very rapid growth in the mixing rate until TDC at which point the combustion chamber begins the expansion and reverse squish flow subprocesses. The former phenomena tends to decrease the eddy angular momentum by corresponding eddy scale growth and decreasing in-cylinder density while reverse squish flow results in mass and momentum exchange which tend to slightly increase the mixing rate. The combination of these two subprocesses is responsible for the sluggish mixing rate growth following TDC. After EOI, expansion and dissipation effects rapidly decrease the turnover rate until EVO. Interestingly, under full load the lower engine speed has a larger peak mixing rate than the higher speed, but the contrary is evident at half load.

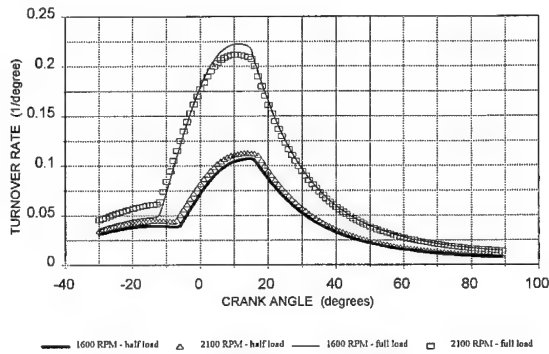


Figure 6 : Fuel-air mixing rate

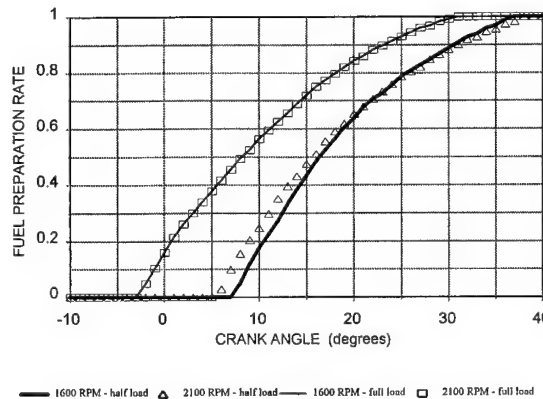


Figure 7 : Normalized Fuel-Air Preparation Rate

For the mixing controlled phase combustion model, the fuel preparation rate is determined from the mixing rate. Basically, injected parcels of fuel are tracked with time and become available to burn once their integrated mixing rate reaches unity. The percent of fuel prepared for combustion under different operating conditions is shown in figure 7. At full load, both engine speeds have the same normalized preparation rate while at half load, the fuel is prepared in a more expedient fashion for the higher speed. This latter behavior at half load is due to a combination of the higher mixing rate at 2100 RPM and the associated later SOI at 1600 RPM.

The calculated fuel burning rates are shown by figure 8. At half load the higher engine speed yields a much larger peak burning rate during the mixing controlled phase. Based on the mixing and preparation rate profiles (figures 6 and 7) as discussed previously, this behavior is reasonable. For full load, the premixed phase peak is larger at lower engine speeds due to the longer ignition delay which allows more fuel to be entrained in the jet shear layer. This directly correlates to higher burning rates and later termination of the premixed phase. The overall peak burning rate, though, is slightly larger for the higher engine speed since more fuel is prepared at a specific time in comparison with the lower speed even though the mixing rate is correspondingly smaller at the higher engine speed (see figure 6).

The experimentally determined gross apparent fuel burning rates are given by figure 9. For these cases, the wall heat loss rate is implicitly lumped in the heat release profile. Thus, lower burning rates are expected as compared to the simulated values which do not explicitly account for wall heat loss. A comparison of the simulated and gross burning rates are shown for 1600 and 2100 RPM in figures 10 and 11. For both speeds, the full load predictions initially perform well but deviate rapidly once the mixing controlled phase begins. At half load, the 1600 RPM prediction shows a much later burning rate profile in comparison to the experimental data while at 2100 RPM there is better agreement. A little deviation at half load is expected in order to account for wall heat loss while a much greater difference is anticipated at full load where higher temperatures and turbulence levels tend to increase thermal losses to the combustion chamber walls. But this does not explain the blatant discrepancy at full load for both engine speeds. Instead, injection rate profile and more accurate SOI appear to be the problem.

A mechanical injection system was employed for fuel delivery in the experimental engine. For this type of system, the SOI and injection profile are controlled by engine speed and fuel within the injector cup. Using cam lift data from a similar engine, a simple injection model was developed to study the effect of these parameters on the

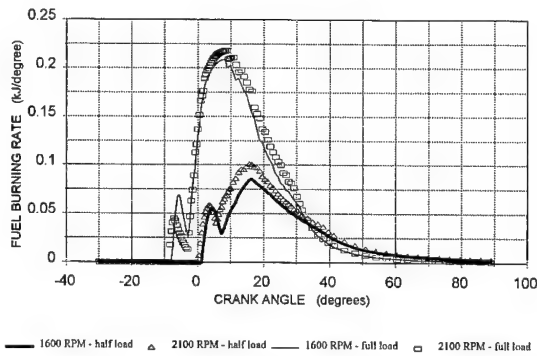


Figure 8 : Simulated net heat release rates

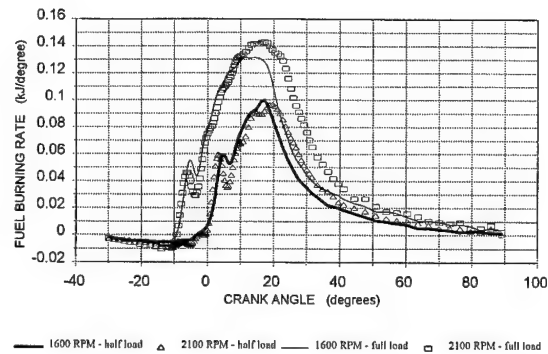
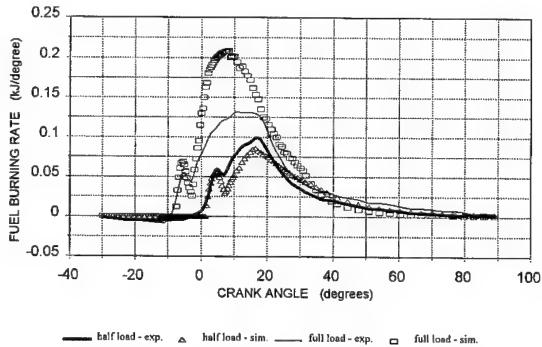
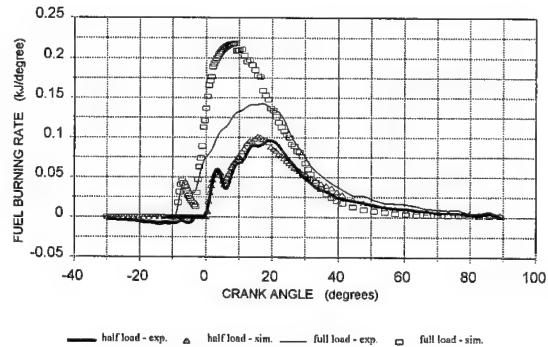


Figure 9 : Experimental gross heat release rates

Figure 10 : Heat release rate comparisons at 1600 RPM
(exp. - experimental, sim. - simulated)Figure 11 : Heat release rate comparisons at 2100 RPM
(exp. - experimental, sim. - simulated)

fueling rate profile. At rated power, the model predicts a trapezoidal profile while with decreasing load and speed, the injection profile becomes more uniform. For the calculated fuel burning rates a square wave was employed for the injection rate and thus, at lower loads and speeds, this assumption appears to be reasonable based on the simple fueling rate model. On the contrary, the full load conditions will be drastically affected with such a uniform injection profile. Intuitively, the trapezoidal profile would provide later fuel preparation and initially a more sluggish mixing rate at SOI which eventually increases in a more rapid fashion during peak injection pressure. This combination of events provides a recipe for a slower increasing and longer fuel burning rate profile throughout the mixing controlled phase. Currently the combustion code does not allow for nonuniform injection profiles, but plans are being made to remedy this situation.

6. Conclusions

A mixing and combustion model has been developed for a low swirl, direct injection diesel. It includes pertinent injection parameters (fuel pressure, cylinder global density, injector nozzle effects, etc.), global in-cylinder thermodynamic conditions, and limited combustion chamber geometry. The model consists of two phases which

account explicitly for fuel burning rate during the entire combustion process. The first phase is a physical representation of the initial flame spread which occurs after autoignition. This approach is based on experimental evidence which supports the existence of a rapidly burning fuel-air shear layer on the injected fuel jet periphery. The second phase is also a characteristic time model which relies on the fuel-air preparation rate. This approach allows the in-cylinder mixing rate (or representative large scale turbulence) to dictate the fuel consumption rate. This trend has also been observed experimentally by many different researchers.

This proposed global combustion model has provided very promising results. For two engine speeds and two load conditions very good agreement was achieved from a magnitude and duration perspective with experimentally determined gross heat release rate profiles. Specifically, the premixed phase submodel performed well under each of the four engine operating points while the mixing controlled phase was more receptive to half load conditions. It is strongly believed that most of the disagreement during the latter combustion phase is strongly a function of the injection rate profile. Intuitively, a trapezoidal fueling rate is expected for the experimental injection system as compared to the uniform profile inherent with the proposed combustion model. The more realistic injection rate would initially result in a more sluggish fuel preparation rate during the entire premixed and early mixing controlled combustion phases, which eventually causes both a lower peak and a more retarded fuel burning rate profile. This effect is more dominant when the injector cup contains more fuel as evident at full load and higher speeds.

7. Recommendations

1. Add a wall heat transfer model to the combustion model in order for more accurate comparison with experimentally determined gross heat release rates. It must include different boundary conditions for each surface in conjunction with geometrical effects on the thermal flow field.
2. Add rate shaping capability to the fuel injection module of the combustion code. Thus, the influence of the fueling rate profile on the overall combustion process may be studied under different engine operating conditions.
3. Pressure data will be obtained for a Detroit Diesel Corporation (DDC) Series 60 engine from the University of Michigan under the U.S. Army Automotive Research Center effort directed by TARDEC. Further analysis will also provide gross heat release rates for comparison and additional validation of the proposed combustion model.
4. Compare the model with experimental data from the ceramic coated piston configuration currently being tested and analyzed at TARDEC. Ideally the baseline and low heat rejection (i.e. ceramic coated) piston should have similar injection rate profiles and thus similar mixing rate profiles. The effect of in-cylinder thermodynamic state on both phases of combustion in conjunction with thermal losses may be studied under very similar global in-cylinder turbulence levels.

References

1. Schihl, P., Bryzik, W., and Atreya, A., "Analysis of Current Spray Penetration Models and Proposal of a Phenomenological Cone Penetration Model", SAE Paper No. 960773, 1996a.
2. Ruiz, F. and Chigier, N., "Parametric Experiments on Liquid Jet Atomization Spray Angle", *Atomization and Sprays*, Vol. 1, pp. 23-45, 1991.
3. Reitz, R.D. and Bracco, F.V., "On the Dependence of Spray Angle and Other Spray Parameters on Nozzle Design and Operating Conditions", SAE Paper No. 790494, 1979.
4. Schihl, P., Bryzik, W., and Atreya, A., "A Large Scale Mixing Model for a Quiescent Chamber Direct Injection Diesel", SAE Paper No. 961040, 1996b.
5. Edwards, C.F., Siebers, D.L., and Hoskin, D.H., "A Study of the Autoignition Process of a Diesel Spray via High Speed Visualization", SAE Paper No. 920108, 1992.
6. Tabaczynski, R., Ferguson, C., and Radhakrishnan, L., "A Turbulent Entrainment Model for Spark-Ignition Engine Combustion", SAE Paper No. 770647, 1977.
7. Schwarz, E., Reid, M., Bryzik, W., and Danielson, E., "Combustion and Performance Characteristics of a Low Heat Rejection Engine", SAE Paper No. 930988, 1993.
8. *Engine Cycle Analysis Operations Manual Version 884*, 1989, available from PEI Consultants, 2020 University Ave., Madison, WI., 53705.

The Fabrication and High-Strain-Rate Properties of Hot-Explosively-Compacted Tungsten-Titanium Alloys

Laszlo J. Kecskes*
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

Ian W. Hall
University of Delaware
Newark, DE 19716-1501

Abstract

A unique, two-step hot explosive compaction (HEC) process has been applied to the tungsten-titanium (W-Ti) system to produce full-density alloys. W and Ti precursor powders of the alloy billet were first preheated by an externally placed combustion synthesis reaction. After a time delay, the powder mixture was isothermal, and was compacted to high density by the pressure waves generated by the detonation of an explosive. The properties of the resultant alloy, consisting of W grains embedded in a Ti-rich, β -Ti/W solid solution, can be adjusted by changes in processing conditions. More importantly, in compressive split Hopkinson pressure bar (SHPB) tests, the 95W-5Ti (wt.%) alloy demonstrates the propensity of the W-Ti alloys to undergo shear localization at high strain rates. The effects of increasing strain rate, orientation, and spatial location within the alloy billet, as well as the sensitivity to specimen geometry were examined. The fabrication methodology, and the results of the high-strain-rate SHPB experiments, as well as the microstructural characteristics of the sheared regions are discussed.

1. Introduction

Uranium (U) alloys (nominally 99.25U-0.75Tiwt.%) are preferred for use in kinetic energy (KE) penetrators, because many of their materials characteristics, such as lower thermal conductivity, lower specific heat, inherently higher flow stress, and higher mass density, give a greater penetration ability than conventional tungsten-nickel-iron heavy alloys (WHA). The increased performance is partly attributed to a novel flow softening behavior, more commonly referred to as adiabatic shear localization (Magness and Farrand, 1990). Reasons for the shearing behavior of U alloys is not well understood. Localization occurs when the rate of thermal softening exceeds that of the rate of work hardening. In ballistic tests with semi-infinite targets, the transformed zones tend to occur at oblique planes with respect to the penetrator-target interface that renders the U alloy penetrator, unlike WHA, able to maintain a "chiseled-nose" shape favorable for enhanced penetration. However, the environmental hazards, primarily the toxicity of U, pose a formidable environmental dilemma associated with their use. Consequently, a new W-based composite alloy was conceived, wherein the desirable properties of U alloys (i.e., increased penetration) and WHA (i.e., nontoxicity) would be combined to develop a new class of high-density, high-strength, and high-hardness KE penetrators.

During the past few years, several adiabatically shearing W-based alloys have become available (Magness, Kapoor, and Dowding, 1995). Typically, the new two-phase composite alloys consist of W grains embedded in a metal matrix, such as titanium (Ti), zirconium, hafnium, or certain types of steels. Model-scale ballistic tests (semi-infinite targets) have demonstrated that the penetration mechanism is modified. The profile of the penetration cavity is similar to that made by the U alloy and different from that made by conventional WHA. Because the new alloys appear to fail by the same mechanism as the U alloys, they may provide a possible alternative for their eventual replacement.

The refractory nature of W and its tendency to form hard, brittle intermetallic phases with some of the metal matrix candidates above, impose limitations on the available routes for alloy fabrication. As a result, the liquid-phase sintering technology developed for the conventional WHA (German, 1985) does not entirely carry over for all of the possible new alloy combinations. This fact was highlighted by several investigators at the 2nd International Conference on Tungsten and Refractory Metals (Bose and Dowding, 1995). Several approaches were explored to produce the new W alloys. In addition to the more traditional liquid-phase and solid-state sintering routes, hot coextrusion (Ohriner et al., 1995), mechanical alloying (Edelman et al., 1995), and alternate fabrication techniques, such as HEC, have been considered. The moderately dense W-Ti system, with its delineating feature of no intermetallics, was most attractive for primary consideration.

A unique, two-step HEC process has been applied to the W-Ti system to produce full-density alloys (Kecskes and Hall, 1995). In this method, referred to as combustion synthesis assisted HEC, or CSA-HEC, hereafter, the W and Ti precursor powders of the alloy billet are first rapidly preheated by an externally placed $\text{Ti} + \text{C} \rightarrow \text{TiC}$ combustion synthesis, or self-propagating high-temperature synthesis (SHS) reaction. After a time delay, the powder mixture, allowed to become isothermal, was compacted to high density by the pressure waves generated by the detonation of an explosive. Unlike liquid-phase sintering or HEC, the temperature cycle in CSA-HEC is short, thereby restricting the complete dissolution of W in Ti. Descriptions of SHS processes are in *Combustion and Plasma Synthesis of High-Temperature Materials* (Munir and Holt, 1990).

In conjunction with a basic assessment of the CSA-HEC process, its limitations, and characterization of the resultant alloy billets, the primary goal of this research effort was to demonstrate the propensity of full-density W-Ti alloys for flow softening behavior (i.e., shear localization) at high strain rates. The effects of increasing strain rate, matrix orientation with respect to the direction of impact, and spatial location within the alloy billet (i.e., periphery or core), as well as the sensitivity to specimen length to diameter ratio (LD^{-1}) were evaluated. The results of these experiments and the microstructural characteristics of the resultant localized regions will be discussed.

2. Experimental Procedure

High purity, 99.9%, 12- μm W and 99.5% pure, -325 mesh (equivalent to -44 μm) Ti, and 99.9% pure, 2- μm graphite powders were used in the fabrication experiments. The W and Ti powders, in a mass ratio of 95:5, were ball-milled for several hours under an argon atmosphere. A molar ratio of 1.0 was maintained for the Ti and graphite powders. The Ti+C mixture was pressed into several doughnut-shaped green compacts.

Pertinent components of the CSA-HEC reaction-compaction assembly design were: the mild-steel fixture; W+Ti powder sample; doughnut-shaped Ti+C green compact; ZrO_2 and graphite sheet insulation; Ti+B igniter mixture; electric match; and the explosive driver package consisting of high-hardness steel anvils, the aluminum piston, and the polyvinylchloride container holding the Amatol (80:20 NH_4NO_3 :TNT) explosive. The explosive driver package, readied separately, was affixed onto the reaction fixture shortly before the initiation of the fabrication process. Prior to compaction, the electric match was activated, and, after the Ti+B mixture ignited the Ti+C reaction, the heat generated by the SHS reaction diffused into the W+Ti powder bed. The outside of the W+Ti bed heated up rapidly with the interior lagging behind. As the periphery began to cool while the interior continued to heat up, at a certain time, the powder bed was isothermal. During this time window (60 to 65 s after ignition), the explosive was detonated, accelerating the piston-anvil to densify the contents of the vessel. After sufficient cooling, the W-Ti billet was extracted. Details of the fabrication process are in Kecskes and Hall (1995).

The uniformity and microstructural properties of the alloy billets were characterized by density measurements, X-ray diffraction (XRD) analysis, optical and scanning electron microscopy (SEM), energy dispersive X-ray spectroscopy (EDS), and microhardness measurements.

Subsequent to the microstructural characterization of the billets, two series of cylindrical samples with LD^{-1} ratios of 1.2 and 1.6 were cut with an electric discharge machine (EDM) from the 95W-5Ti alloys, and have been compression tested at strain rates of up to 10^3 s^{-1} using the SHPB apparatus at the University of Delaware. The experimental and data reduction procedures are discussed at length by Frey (1990). For comparison, baseline experiments were also conducted at a low, 10^{-2} s^{-1} , quasi-static strain rate.

3. Results and Discussion

After an initial process development and optimization, the experimental program consisted of the characterization of the CSA-HEC billet with a concurrent investigation of the shear localization behavior of the 95W-5Ti (wt.%) alloy at quasi-static and high-strain rates.

3.1 CSA-HEC Process Development and Optimization

Preceding the compaction experiments, the temperature of the powder bed during the preheating phase was measured. Results showed that the molar ratio of the exothermic material to the sample significantly affected the heating rate, magnitude, and uniformity of the temperature within the W+Ti powder bed. This information was also necessary to identify the appropriate point in time for consolidation. With a most suitable molar ratio of 6.3, the temperature at the periphery was in excess of 2,000°C, whereas in the interior it did not exceed the melting point of Ti ($T_m \text{ Ti} = 1,670^\circ\text{C}$). In SEM and EDS examination of the heated W+Ti sponge, the Ti was found to coat the nearby W grains, thereby permitting the limited dissolution of W into Ti to form a β -Ti/W solid solution.

Optimization of the process was completed with experiments wherein the mass of the explosive charge was varied. It was found that the detonation of smaller amounts of explosive imparted insufficient energy to densify the W+Ti powder mixture. In contrast, coupled with a longer shock wave duration, the detonation of a large amount of explosive imparted too much energy, thereby damaging the sample. Therefore, the amount was adjusted to maximize densities without significant shock-rarefaction-wave-induced cracking and delaminations in the samples.

3.2 W-Ti Alloy Characteristics

The 95W-5Ti alloy, with a theoretical density (TD) of 16.55 gcm^{-3} , was successfully compacted to full density (i.e., 98. + % TD). The alloy, consisting of W grains embedded in a partially continuous β -Ti/W matrix, had an average Vickers microhardness of $5.0 \pm 0.1 \text{ GPa}$. This value is similar to that of mild steel. XRD analysis revealed the body centered cubic peaks of W and matrix phases only. When compared with the W precursor, the W grains in the alloy showed little distortion. The composition of the β -Ti/W phase was estimated at 40W-60Ti.

The billet could be subdivided into an annular periphery and an inner core. The edge region tended to contain horizontal and conical cracks. In contrast, the core was usually free of cracks. At the core, the composite consisted of W grain aggregates embedded in a generally discontinuous, preferentially oriented, Ti-rich, β -Ti/W matrix. In the periphery, the matrix also contained features ranging from fine in-grain and fine intragranular precipitates and duplex, monotectoid structures. The discontinuity or incomplete dispersion of Ti was brought about by the considerable size difference of the precursors, their relative melting points, and temperature at the time of compaction. Parallel and perpendicular sections (with respect to the compaction axis) of the billet revealed that the distribution of the phases was isotropic in the perpendicular cut; however, the phases were elongated in the parallel cut. This orientation effect simply arose during the uniaxial collapse of the initially isotropic W+Ti powder bed. The spatial dependence of the matrix substructure and its heterogeneity, especially within the periphery, was found to be determined by the precompaction temperature history of the W+Ti powder bed.

Further experiments were centered on the ability to alter or improve the existing alloy microstructure. The use of different sized precursors resulted only in minor or little change in alloy structure. Densification improved with a finer Ti precursor ($10 \mu\text{m}$), although at the expense of a more fibrous substructure. The alloy was less oriented with a larger W precursor ($\sim 325 \text{ mesh}$), but, in turn, the billet was less dense. The effect of W:Ti stoichiometry on the dispersibility of the W grains in the alloy was significant. In the most promising, alternate, proof-of-principle approach, the W precursor was coated with a thin layer of Ti and the coated powder compacted. Though not fully dense, the resultant alloy had a density approaching that of W and consisted of W grains isolated by a thin layer of Ti-rich matrix. Because of limitations of the physical vapor deposition system (i.e., the purity of argon), these samples tended to be more brittle.

3.3 High-Strain-Rate Properties of the 95W-5Ti Alloy

In the SHPB tests, both intrinsic and extrinsic factors influenced the formation of shear instabilities within

the 95W-5Ti alloy. The ultimate stress levels exhibited by the samples were independent of the strain rate. As shown in Figure 1, the plot of ultimate stress vs. ultimate strain for an LD^{-1} of 1.2 illustrates the shear localization susceptibility of the 95W-5Ti alloy. The three sets of data, two perpendicular and one parallel orientation, are similar, showing a slight increasing strain sensitivity of the ultimate stress. The plot is divided by a horizontal line marking the transition from extrinsic to intrinsic behavior. The vertical band at a strain of 0.12 ± 0.01 demarcates the shear localization threshold. Samples 6, 10, and 34 above the horizontal line, but below the shear instability threshold were barreled, whereas samples 39, 7, and 38 above the threshold failed via shear localization. Samples within the threshold band were barreled (9), contained localizations that did not propagate (29, 41, and 44), or sheared (33 and 43).

Failure generally occurred along surfaces of maximum shear (i.e., 45° cones, or body diagonals). Surfaces of sample fragments revealed the presence of both brittle, tensile inter- and transgranular cleavage, and ductile, shear failure. Above the horizontal line, sample 40 failed by shear, while in sample 28 localization developed, but did not propagate. Below the horizontal line, samples failed at much lower shear thresholds, indicative of an extrinsic, secondary instability for the onset of flow-softening behavior. Note, all samples below this line, except 8, which barreled, have failed via shear (42, 37, 31, 30, and 24), or shear coupled with extensive fragmentation (26 and 27). Although localization and shear occurs, in these cases the driving mechanism is attributed to a variety of impact surface and bulk material defects of the samples (e.g., pits, cracks, or grooves) that act as stress raisers. For instance, the lower ultimate stresses of sample 42, 37, and 24 are attributed to a deep groove on their impact surface produced by the EDM cutting tool. Similarly, the premature failures of samples 30, 31, 26, and 27 are likely associated with chipped and fractured corners at the impact surface introduced during machining.

Microstructural analysis of the barreled sample cross sections revealed either no localization, or localized regions that failed to propagate. As shown in Figure 2, cross sections of recovered, sheared fragments exhibit very narrow widths for the localized zone ($50 \mu\text{m}$) and very large (50-100:1) W grain elongations within the zone. The failure mode was essentially independent of the β -Ti/W matrix's spatial location and only slightly dependent on its orientation. The elongated W grain morphology in the localized region was identical for both orientations. However, whereas in the perpendicular orientation the W grains rotated only 45° into the shear band, in the parallel orientation, the W grains underwent a much greater rotation of 135° . Finally, while the underlying nature of the thermomechanical instabilities and flow softening behavior of samples with LD^{-1} of 1.6 did not differ from those with LD^{-1} of 1.2, the localization tended to occur at a lower threshold strain of 0.09. That is, as expected, the effect of greater friction between the anvils and the sample was only reflected in a lower propagation threshold.

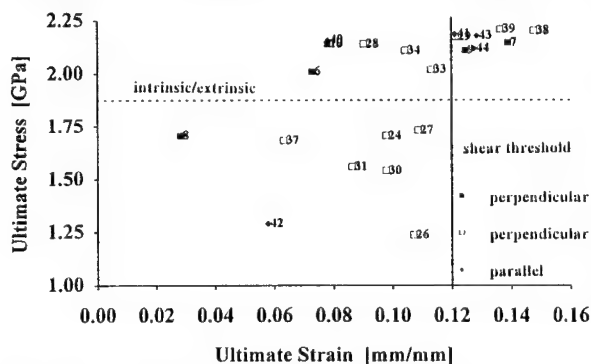


Figure 1. Failure strength and strain sensitivity of the 95W-5Ti alloy at high strain rates.

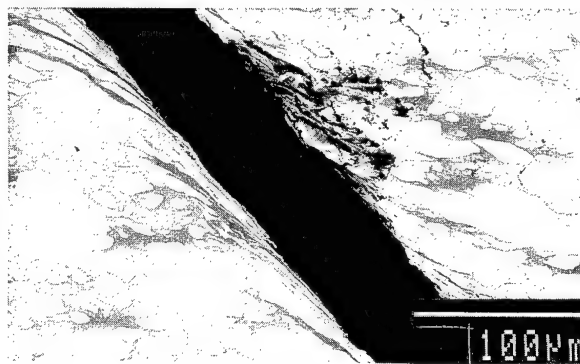


Figure 2. Typical shear localization structure in the failed 95W-5Ti alloy samples.

In the quasi-static compression tests, the 95W-5Ti alloy was found to be more ductile than at high strain rates and exhibited spatial location sensitivity. Specifically, the sample from the core of the billet underwent gross plastic deformation with a strain of about 0.15. In addition to an overall compressive deformation of both the W grains and the β -Ti/W matrix, limited localization developed along the body diagonals of the cylinder. The sheared zone was considerably wider and diffused with W grain elongations of about 5:1. In contrast, the sample taken from the edge of the billet did not support the load and failed at a strain of 0.03. In fact, the sample was so brittle that it could not be sectioned for cross-sectional examination.

4. Conclusions

Application of a CSA-HEC method to the W-Ti system yields two phase composites consisting of W grains embedded in a β -Ti/W solid solution matrix. The alloy is oriented, heterogeneous, and segregated; however, its structure can be modified with relative ease. More importantly, the 95W-5Ti alloy exhibits the desired flow softening behavior for the purpose it was designed. Results of the SHPB tests indicated that both intrinsic and extrinsic factors contributed to the thermomechanical instabilities in the alloy. The ultimate stress of the alloy was strain-rate independent and was only slightly dependent on strain. The shear localization susceptibility was found to occur above a certain threshold strain level. Failed surfaces indicated competing modes of a generally brittle, tensile inter- and transgranular fracture, and ductile, shear localization failure. Quasi-static compression results were consistent with the segregated structure of the billet.

Acknowledgments

The authors would like to extend their thanks to Mr. Frederick Pierce for his assistance with the explosive operations, fixture preparation, and sample analysis and to Dr. Michael Staker for his valuable suggestions in the course of the high-strain-rate testing.

References

1. Magness, L.S. and Farrand, T.G., "Deformation Behavior and Its Relationship to the Penetration Performance of High-Density KE Penetrator Materials," *Proc. 17th Army Science Conf.*, Durham, NC, 1990, Army Science Board, Washington, DC, Vol. 2, pp. 149-164.
2. Magness, L.S., Kapoor, D., and Dowding, R.J. "Novel Flow-Softening and Flow-Anisotropy Approaches to Developing Improved Tungsten Kinetic Energy Penetrator Materials," *Mater. and Manuf. Processes*, Vol. 10, No. 3, pp. 531-540, May 1995.
3. Bose, A. and Dowding, R.J. (eds.), *Tungsten and Refractory Metals - 1994*, Metal Powder Industries Federation, Princeton, NJ, 1995.
4. Ohriner, E.K., Sikka, V.K., and Kapoor, D., "Processing and Properties of Extruded Tungsten-Hafnium and Tungsten-Steel Composites," *Tungsten and Refractory Metals - 1994*, Bose, A. and Dowding, R.J. (eds.), Metal Powder Industries Federation, Princeton, NJ, pp. 219-226, 1995.
5. Edelman, D.G., Pletka, B.J., and Subash, G., "Mechanical Alloying of W-Hf-Ti Alloys," *Tungsten and Refractory Metals - 1994*, Bose, A. and Dowding, R.J. (eds.), Metal Powder Industries Federation, Princeton, NJ, pp. 227-234, 1995.
6. German, R.M., *Liquid Phase Sintering*, Plenum Press, New York, NY, 1985.
7. Kecskes, L.J. and Hall, I.W., *Hot Explosive Consolidation of W-Ti Alloys*, ARL-TR-669, Army Research Laboratory, Aberdeen Proving Ground, MD, January 1995.
8. Munir, Z.A. and Holt, J.B. (eds.), *Combustion and Plasma Synthesis of High-Temperature Materials*, VCH Publishers, Inc., New York, NY, 1990.
9. Frey, T.J., *High Strain Rate Effects on Composite Material Properties*, M.S. Thesis, Dept. of Mechanical Engineering, University of Delaware, Newark, DE, 1990.

Concrete Constitutive Modelling in High Velocity Penetration Analysis

V. M. Gold*, G. C. Vradis**, and J. C. Pearson*

*US Army ARDEC
AMSTA-AR-AEE-W, Bldg. 3022
Energetics and Warheads Division
Picatinny Arsenal, N. J. 07806-5000

**Department of Mechanical, Aerospace and Manufacturing Engineering
Polytechnic University
Six Metrotech Center, Brooklyn, N.Y. 11201

Abstract

A combined numerical and analytical study of penetration of concrete by high velocity ($\sim 1.8 \text{ km/s}$) projectiles has been conducted. The effects of concrete's constitutive modelling on penetration calculations were studied and are discussed. The results of the analysis are compared with the available experimental data. All constitutive models studied accounted for the compressibility of concrete. Regardless of the value of the yield strength, application of a constant yield strength model resulted in nearly identical hole profiles which, in all cases, significantly disagreed with the experimental data. Crater profiles calculated with the pressure-dependent yield model showed good agreement with the available experimental data. The reduction in the calculated crater profiles was attributed to the increased target resistance to penetration. The paper presents a detailed analysis which shows that inclusion of the yield strength-pressure dependency in the constitutive model significantly increases the target's resistance to penetration.

1. The Numerical Model

In the present work, the numerical experiments have concentrated on modelling a representative experiment with a 19cm long and 1.3cm diameter copper projectile impacting a plain concrete target with velocity $V_0 = 0.1836 \text{ cm}/\mu\text{s}$ (test PA8Cu1.4; see Gold *et al.*, 1996a). The numerical analysis presented here was performed using the CALE (Tipton, 1991a) computer program. In modelling the constitutive response of concrete targets, all models studied accounted for compressibility of concrete, which was modeled employing the porous equation of state model. Detailed description of this model and its implementation in CALE can be found in Tipton (1991b).

CALE's procedure for calculation of the elastic-plastic flow employs the von Mises yielding criterion. The yielding behavior of concrete was modeled using the following models: (1) constant yield-strength model, $J_2 = k = \text{Const}$ and (2) pressure dependent yield-strength model, $J_2 = k(p)$, where J_2 is the second deviatoric stress invariant and k or $k(p)$ are either an empirical strength constant, or an empirical function (of pressure p), respectively. A detailed description of the governing equations, the boundary conditions, and the computational mesh employed in the model can be found in Gold *et al.* (1996b).

2. Results of the Numerical Analysis

The analysis presented in this work addresses a number of issues pertaining to constitutive modelling of concrete in high velocity penetration analysis. Two concrete yield-strength models were studied: (1) the constant yield-strength model and (2) the pressure dependent yield-strength model. Hole profiles resulting from the numerical experimentation with these models are shown in Fig. 1 (a-b). In the case of the constant yield model, the disagreement between the calculated hole profiles and the experimental data was rather significant (see Fig. 1 (a)), regardless of the value of the yield strength parameter k . In the case of the pressure dependent yield-strength model, the analysis resulted in hole predictions, which not only reproduced the experimental depth of penetration, but also matched the measurements of the profile of the tunnel portion of the crater (see Fig. 1 (b)). Thus, before one even attempts to understand the relationship between the target constitutive modeling and the target resistance to penetration, one has to identify principal factors which determine the capability of the target material to resist penetration.

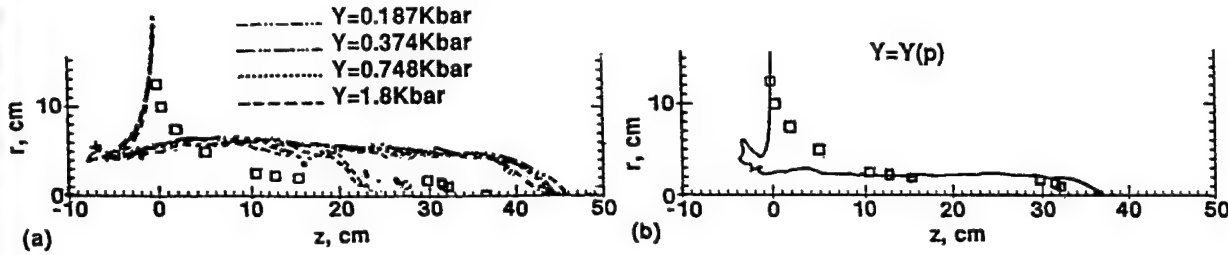


Fig. 1. Resulting hole profiles employing (a) the constant yield-strength model and (b) the pressure dependent yield-strength model. \square Experimental data Gold *et al.* (1996a).

It can be hypothesized that one of the principal parameters that affect the target resistance to penetration is the pressure at the target/penetrator interface. Tate (1967) has shown that, if there exists a difference in the penetration characteristics, this difference must manifest itself during the stage of the steady-state penetration. During the quasi steady-state penetration, the motion of the penetrator can be viewed within the framework of the one-dimensional modified hydrodynamic theory of penetration (Tate, 1967). According to this theory, the pressure at the target/penetrator interface is given by the following:

$$\frac{1}{2}\rho_{p0}(v-u)^2 + Y_p = \frac{1}{2}\rho_{t0}u^2 + R_t \quad (1)$$

where ρ_{p0} and ρ_{t0} are the densities of the penetrator and the target materials, respectively. The velocity of the undeformed portion of the penetrator is v , u is the penetration rate, and Y_p and R_t are empirical parameters relating to the strengths of penetrator and target materials, respectively.

The empirical yield strength function $Y=Y(p)$ which was employed in the analysis using the pressure dependent yield model, is that of Smith (1993). During the quasi steady-state penetration the average pressure at the target/penetrator interface is approximately 0.02 Mbar. The corresponding averaged yield strength is approximately 1.6 Kbar. Intuitively, one would expect that the higher values of the "average" strength have to result in the higher values of the pressures. And yet, the calculations show that in the case of the pressure dependent yield model (the "average" yield is approximately 1.6 Kbar), the time averaged pressure p^* is higher than the pressure resulting from the 1.8 Kbar constant yield model. It follows then, that the target resistance to penetration is not related simply either to the pressure, or to the yield strength alone. In order to identify the principal factors affecting the target resistance to penetration, one has to examine the structure of the plastic flow field in the target.

The structure of the elastic-plastic flow field in the target can be examined from the results presented in Fig. 2. Isolines of the equivalent plastic strain shown in Fig. 2. represent the approximate position of the elastic-plastic boundary in the target. The equivalent plastic strain is defined as $\gamma = (2/3)\sqrt{(\epsilon_{1p}-\epsilon_{2p})^2 + (\epsilon_{2p}-\epsilon_{3p})^2 + (\epsilon_{3p}-\epsilon_{1p})^2}$, where ϵ_{1p} , ϵ_{2p} , and ϵ_{3p} are the principal plastic strain components. It is interesting to note that all numerical experiments resulted in isolines with shapes that are approximately hemispherical, regardless of the employed target constitutive model. The hemispherical symmetry in the target deformation field appears to be

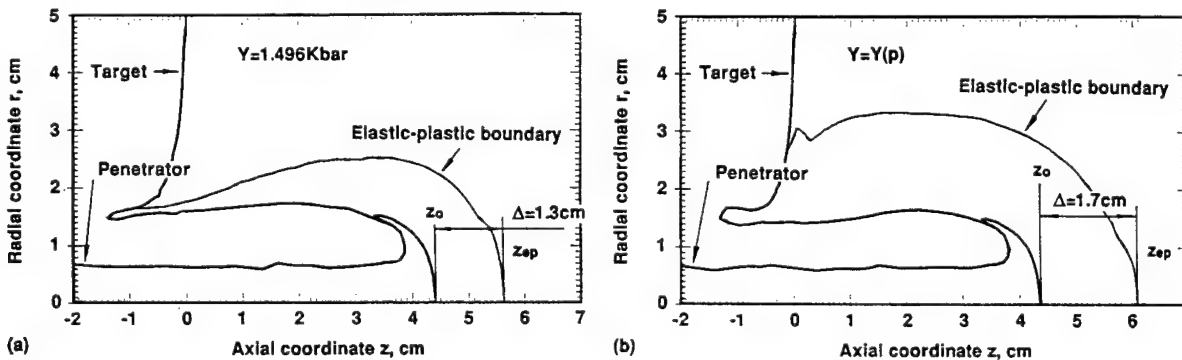


Fig. 2. Structure of the elastic-plastic flow field in target (time $t=35\mu s$, $\gamma=0.062$).

an intrinsic characteristic of the high velocity penetration process proper, rather than a result of a particular type of constitutive model.

The numerical experiments showed that, depending on the particular structure of the employed constitutive model, the size of the plastic zone Δ ($\Delta = z_{ep} - z_0$, see Fig. 2) was either significantly affected by the model, or the effect was negligible. In particular, all numerical experiments with the constant yield-strength model resulted in approximately the same value of $\Delta \approx 1.3 \text{ cm}$, regardless of the value of the yield strength parameter Y . In contrast, for pressure dependent yield-strength model, the numerical experiments resulted in a significant 30% increase in the size of the plastic zone ($\Delta \approx 1.7 \text{ cm}$).

In the above approach, the target resistance to penetration is associated with the amount of energy per unit volume dissipated in the plastic flow. If the target is an ideal fluid, then, according to the ideal hydrodynamic theory of penetration the penetration resistance of the target is a function of the target density only. If the target is a solid with a finite strength, then one could argue that the stronger the target, the larger the value of the energy dissipation per unit volume must be. In contradiction to this intuitive conclusion, the numerical experiments indicated the opposite: the higher the target resistance to penetration, the larger the size of the plastic zone Δ , and, therefore, the lesser the amount of the dissipated energy per unit volume. The explanation of these "controversial" results needs a detailed analysis of the notion of the "target resistance to penetration".

3. The Target Resistance to Penetration: Analysis of the Axial Momentum Balance Equation on the Line of Axial Symmetry

Assuming that the penetration is approximately steady state, Pidsley (1984) showed that on the center line, $r=0$, the axial momentum balance equation can be reduced to the following equation:

$$\rho v_z \frac{dv_z}{dz} = \frac{d}{dz}(s_{zz} - p) + 2 \left. \frac{\partial s_{rz}}{\partial r} \right|_{r=0} \quad r=0, z_0 - l \leq z \leq z_{ep} \quad (2)$$

where l is the length of the penetrator. Considering an arbitrary domain in the target $z_0 \leq z \leq z^*$ ($z^* \leq z_{ep}$), in which $\rho = \rho_t(z)$, the pressure at the target/penetrator interface can be found by integrating Eq. (2) over the interval $[z_0, z^*]$, resulting in an expression similar to one obtained by Batra and Wright (1986):

$$p(z_0) = \underbrace{\int_{z_0}^{z^*} \rho_t v_z dv_z + p(z^*) + s_{zz}(z_0) - s_{zz}(z^*)}_{R_h} - 2 \int_{z_0}^{z^*} \left. \frac{\partial s_{rz}}{\partial r} \right|_{r=0} dz \quad (3)$$

The term R_h appearing on the right hand side of Eq. (3) represents the "hydrodynamic" component of the pressure at the target/penetrator interface. In this work, concrete was modelled as porous, or "crushable" material. Numerical experiments showed (see Gold *et al.*, 1996b) that once the concrete is crushed and the porosity φ is "squeezed out", the changes in the density of the crushed concrete become negligible. Thus, assuming that the coordinate z^* is within this region of completely crushed concrete (with $\rho_t \approx \text{const}$), the term R_h of Eq. (3) can be integrated, which results in the following:

$$p(z_0) \approx \underbrace{\frac{1}{2} \rho_t v_z^2(z^*) - \frac{1}{2} \rho_t v_z^2(z_0) + p(z^*) + s_{zz}(z_0) - s_{zz}(z^*)}_{\Delta s_{zz}} - 2 \underbrace{\int_{z_0}^{z^*} \left. \frac{\partial s_{rz}}{\partial r} \right|_{r=0} dz}_{R_s} \quad (4)$$

In the case of an axially symmetrical plastic flow, a closed form analytical solution for the deviatoric stress s_{zz} can be found on the center line $r=0$. In particular, if the material is incompressible and the yield condition is given by the von Mises yield criterion, one can show that at the center line $r=0$, $s_{zz} = 2k/\sqrt{3}$. Thus, in the case of the constant yield-strength model, $\Delta s_{zz} = 0$. In the case of the pressure dependent yield-strength model, $k(p) = Y(p)/\sqrt{3}$, $\Delta s_{zz} = (2/3) \cdot [Y(p(z_0)) - Y(p(z^*))]$. Assuming that above the compaction pressure limit p_c , the strain-hardening of concrete is insignificant (Smith, 1993), the last equation results also in $\Delta s_{zz} \approx 0$.

Comparison between the relative magnitudes of the terms of Eq. (4) (see Fig. 3) shows that the terms that contribute the most to the target/penetrator interface pressure are the pressure term $p(z^*)$ and the

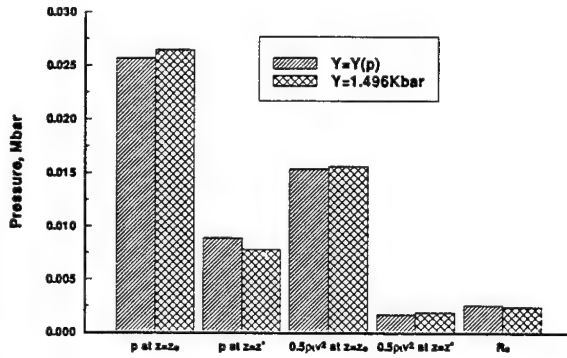


Fig. 3. Comparison between the instantaneous values of various terms comprising the pressure at the target/penetrator interface ($t=35\mu s$).

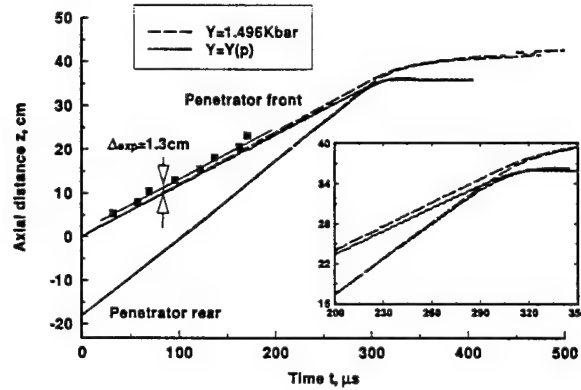


Fig. 4. Trajectories of the front and the rear of the penetrator for the constant yield-strength and the pressure dependent yield-strength models. ■ Experimental data Gold *et al.* (1996b)

"hydrodynamic" term, $0.5\rho_t[u^2(z^*)-v^2(z_0)]$. During the high velocity penetration, the pressure in the target material fluctuates rapidly and within rather wide limits. These rapid and large fluctuations prohibit establishing a direct relationship between the target constitutive modelling and the pressure terms $p(z_0)$ and $p(z^*)$, unless the instantaneous values of $p(z_0)$ and $p(z^*)$ (shown in Fig. 3) are time averaged. In contrast to the pressure term $p(z^*)$, the "hydrodynamic" term (which represents the kinetic energy of the target material), is not susceptible either to large, or to rapid fluctuations, which makes it a convenient parameter for examining effects of the target constitutive modeling.

The calculated trajectories of the front and the rear of the penetrator for the constant yield-strength and the pressure dependent yield-strength models are shown in Fig. 4. The slopes of the trajectories of the penetrator front and the rear represent the penetration rate u and the penetrator rigid body velocity v , respectively. According to the modified hydrodynamic theory of penetration, the penetrator erosion rate is given by $dl/dt=-(v-u)$, where l is the instantaneous length of the penetrator. Numerical experiments showed that, after the concrete constitutive model was changed to the pressure-dependent yield-strength model, the penetration rate u was decreased. Since the rigid body velocity v remained unchanged, the absolute value of the penetrator erosion rate $|dl/dt|$ was increased proportionally. According to the hydrodynamic theory of penetration, the penetration is accomplished at the expense of the penetrator material. Therefore, the faster the penetrator erodes, the faster it spends the available kinetic energy which, finally, results in a decrease in the penetration. This is precisely the reason why the pressure-dependent yield-strength model resulted in a decrease in the penetration depth. Thus, it is reasonable to relate the notion of target resistance to penetration not to the pressure at the target/penetrator interface $p(z_0)$, and not to strength related parameters, such as R_s or Δs_{zz} , but rather to the penetration rate u .

Since we have established that the target resistance to penetration is directly related to the penetration rate u , it would be instructive to examine the structure of the term R_h . Since R_h is identical with $\rho_t u^2/2$ of Eq. (1), rearranging terms of Eq. (3), results in the following equation:

$$R_h = \frac{1}{2}\rho_t u^2 = \underbrace{p(z_0) - p(z^*)}_{\Delta p} + \underbrace{s_{zz}(z^*) - s_{zz}(z_0)}_{-\Delta s_{zz}} + \underbrace{2 \int_{z_0}^{z^*} \frac{\partial s_{rz}}{\partial r} dz}_{-R_s} \quad (5)$$

Since changes in the target constitutive modelling have a negligible effect on the terms Δs_{zz} and R_s , it follows then that changes in the penetration rate u are affected mostly by changes in the pressure gradient Δp . Thus, the target resistance to penetration should increase with decreases in the pressure gradient Δp .

Fig. 5 shows calculated profiles of the pressure p along the center line $r=0$, at time $t=35\mu s$ after impact. The shapes of the pressure profiles show that decreases in the slopes dp/dz correlate very well with increases in the target resistance to penetration. Thus, the results of the numerical experiments are in complete agreement with the predictions of Eq. (5), which states that the penetration rate is proportional to the pressure gradient Δp .

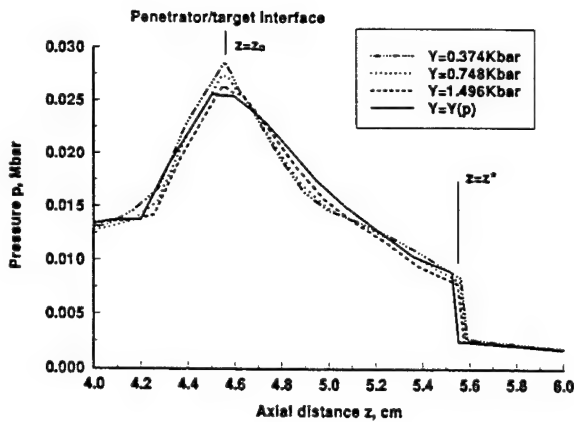


Fig. 5. Calculated profiles of pressure p along the center line $r=0$ ($t=35\mu s$).

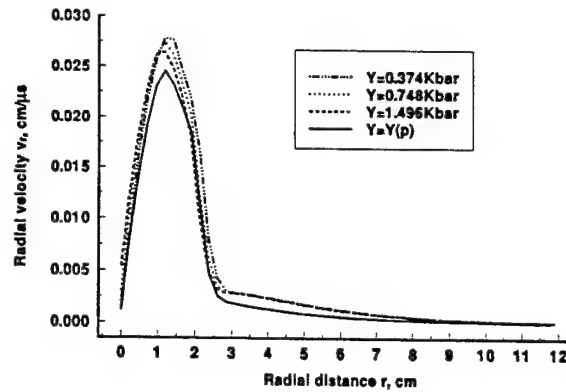


Fig. 6. Profiles of the radial velocity component v_r versus radial coordinate r ($z=5.2\text{ cm}$, $t=35\mu s$).

This general tendency, that the gradients of plastic flow field variables consistently diminish with increases in the target resistance to penetration, is related to increases in the size of the plastic zone discussed in the preceding section and shown in Fig. 2. The nature of the relationship between the gradients of the flow field variables and the extent of the plastic zone is as follows. Numerical experiments show that the changes in the target constitutive modelling examined here do not significantly affect plastic flow field variables at either the target/penetrator interface z_0 , or at the elastic-plastic boundary z_{ep} . It follows then that since inside the plastic zone $z_0 \leq z \leq z_{ep}$ flow field variables change monotonically, once the extent of the plastic zone ($\Delta = z_{ep} - z_0$) is increased, the gradients of flow field variables, $dg(z)/dz \sim \Delta g(z)/\Delta$, are decreased proportionally to increases in Δ .

Another important penetration parameter that is significantly affected by the target constitutive modelling is the radial component of the velocity, v_r . Profiles of v_r at a distance of approximately 0.7 cm in the front of the target/penetrator interface, at $t=35\mu s$ after impact, are shown in Fig. 6. Comparison between these curves shows that with increases in the target resistance to penetration, the radial component of the velocity is consistently decreasing. The relationship between the target resistance to penetration and the final crater diameters is as follows. Behind the penetrator mushroom, the hole continues to grow under the inertial forces that are due to the radial motion of the target material. The magnitude of these forces is determined by the intensity of the target material flow at the penetrator mushroom. The smaller the radial component of the velocity is, the smaller are the inertial forces and, therefore, the smaller the resulting crater diameters are.

Another important parameter that determines the final hole diameter is the radial extent of the plastic zone (adjacent to the surface of the hole) behind the penetrator mushroom. Numerical experiments showed (see Fig. 2) that, when the target constitutive model was changed to the pressure-dependent yield-strength model, the extent of the plastic zone in the target was increased not only in front of the penetrator, but also radially, behind the penetrator mushroom. Increases in the plastic zone size increase the dissipation of the energy (through which the plastic work is converted to heat) stored in the expanding hole. Thus, the increase in the plastic zone size combined with the decrease in the radial velocity profile resulted in a significant decrease in the final crater diameter shown in Fig. 1.

References

- (1) Batra R. C. and Wright T. W., 1986, "Steady State Penetration of Rigid Perfectly Plastic Targets", *Int. J. Engng. Sci.*, Vol. 24, pp. 41-54.
- (2) Gold V. M., Vradis G. C., Pearson J. C., 1996a, "Penetration of Concrete by Eroding Projectiles: Experiments and Analysis", *ASCE J. Eng. Mech.*, Vol. 122, No. 2, pp. 145-152.
- (3) Gold V. M., Vradis G. C., Pearson J. C., 1996b, "A Study of Constitutive Models for Concrete Penetration Analysis", *ASCE J. Eng. Mech.*, Vol. 122, No. 3, pp. 230-238.
- (4) Pidsley P. H., 1984, "A Numerical Study of Long Rod Impact onto a Large Target", *Int. J. mech. Sci.*, Vol. 32, pp. 315-333.
- (5) Smith, J., 1993, Private Communications, Lawrence Livermore National Laboratory, Livermore, California.
- (6) Tate A., 1967, "A Theory for Deceleration of Long Rods After Impact", *J. Mech. Phys. Solids*, Vol. 15, pp. 387-399.
- (7) Tipton R. E., 1991a, "CALE Users Manual", Version 910201; 1991b, "Porous EOS in CALE Used to Model Concrete", Lawrence Livermore National Laboratory, Livermore, California.

Source Characterization Modeling for Demil Operations

C. A. Biltoft*
West Desert Test Center
U.S. Army Dugway Proving Ground
Dugway, UT 84022-5000

E. S. Oran, J. P. Boris, and C. A. Lind
Naval Research Laboratory
Washington, DC 20375

W. J. Mitchell
U.S. EPA Office of Research and Development
Research Triangle Park, NC 27711

Abstract

Reactive flow computational fluid dynamics (CFD) is being applied to the elimination of unwanted conventional munition inventories. This modeling, which includes numerical time-resolved compressible flow and detonation chemistry algorithms, aids in the assessment of burn/detonation emissions factors, evaluation of mitigation strategies, and design of open burn and open detonation containment facilities. A measurement and modeling program is being implemented to guide the elimination of the unwanted inventories in a safe, cost-effective, and environmentally responsible manner.

1. Introduction

During the Cold War the United States of America became the "Arsenal of Democracy," accumulating stockpiles of conventional munitions, propellants, and related materials in hundreds of depots at home and abroad. Most of this arsenal remains unexpended. The post Cold War military drawdown has thus left the United States with large demilitarization (demil) inventories of obsolete energetic materials and explosive devices. This inventory is estimated to be in excess of 400,000 tons, accumulating at a rate of 40,000 to 50,000 tons per year. Base realignment and closure (BRAC) is slated for many of the arsenals and depots housing this material, increasing the urgency for finding safe, inexpensive, and environmentally acceptable disposal methods.

The mass of the demil inventory consists of roughly 20 percent small munitions (below 40 mm) and 80 percent large munitions (rockets, tank and artillery rounds). A traditional disposal method has been the unregulated open burn (OB) or open detonation (OD) in open pits. However, noise, safety, and environmental concerns have curtailed OB and OD, stimulating a search for alternative disposal techniques. Small munitions can be incinerated in rotary kilns, but large ones cannot be disposed of in this manner. Some of the inventory can be synthesized into reusable materials or can be rendered harmless using novel chemical or biological processes. However, these alternative disposal technologies do not address the magnitude and scope of the demil problem, entail significant expense, or present environmental hazards in the form of processing byproducts. Also, some of the inventory is too dangerous for movement to processing facilities.

Thus, cost and safety issues force the international demil community to re-examine large-scale OB/OD as a means to eliminate the bulk of the demil inventory.

Although a few small-scale demil operations are presently permitted, large scale OB/OD awaits documentation demonstrating its safety and efficacy. Dugway Proving Ground (DPG), the Naval Research Laboratory (NRL), the Office of Research and Development of the U.S. Environmental Protection Agency (EPA), and the Lawrence Livermore National Laboratory (LLNL) are cooperating in a Strategic Environmental Research and Development Program (SERDP) study. A major SERDP goal is to devise, test, and implement procedures that will meet the Nation's demil needs in a cost-effective, safe, and environmentally responsible manner. Our SERDP objective is to define the meteorological, topographical, noise mitigation, and safety constraints under which large scale OB/OD activities can be conducted. Key components of this project are the detailed documentation of burn and detonation products, the use of CFD codes to characterize OB/OD events, and atmospheric dispersion modeling for the materials released into the atmosphere.

2. The Measurement Program

Although open-air burns and detonations have been used for demil inventory disposal for over 150 years, most disposal operation institutional knowledge is qualitative, based on individual perceptions of how well various disposal techniques work. While providing valuable insights, qualitative understanding cannot be extrapolated to types and scales of activities that have not been attempted, nor does it provide the information base needed to satisfy regulatory concerns. The need to evaluate new disposal techniques and the requirement for documentation acceptable to regulatory agencies provides impetus to a products-of-burn/detonation measurement program.

Beginning in 1989, the U.S. Army began products of detonation and combustion measurements for various explosives and propellants. Preliminary tests showed that four classes of compounds could be measured: (1) certain gases (CO_2 , CO , NO_x , O_3 , HCN), (2) particles, (3) volatile organic compounds, and (4) semi-volatile compounds. Small scale experiments, many of which have been conducted in the DPG "BangBox" (an enclosed combustion/detonation facility), have demonstrated the feasibility of quantifying emissions factors, including the decay rates of primary and secondary combustion/detonation products. Comparisons with outdoor experiments have also shown that the emissions factors determined from the BangBox are equivalent to those obtained from detonations and burns conducted in the open air (Johnson, 1992). Limited OB/OD operations are now permitted based on BangBox results, but expanded demil operations require scaling to larger and more complex scenarios and the conversion of emissions factors into atmospheric dispersion model source terms. Consequently, we are applying reactive flow CFD to large-scale detonation facility designs, mitigation modeling, and emissions factors/dispersion model source term conversion.

3. Computational Fluid Dynamics Codes

Present modeling of detonation processes is based on empirically-derived parameterizations, which are largely devoid of event physics, cannot depict time-dependent features, and are too limited in geometry to adequately characterize many OB/OD events. Because the open burn or detonation of energetic materials produces a complex array of products which are highly dependent upon initiation conditions, characterization of these events is best accomplished using reactive-

flow, thermochemical, and gas kinetic CFD codes. Salient components of CFD codes are two- or three-dimensional grid generators, a flow solver with user-selectable physics and reactive chemistry algorithms, and a graphics package that permits visualization of time-dependent events. Oran and Boris (1993) provide a comprehensive review of reactive flow numerical simulations.

FAST3D and CHEETAH are examples of CFD codes with OB/OD applications. Developed at NRL, FAST3D is based upon the flux-corrected transport (FCT) algorithm with complex geometry provided by the virtual cell embedding algorithm (Boris et al., 1993). The FCT algorithm (Boris and Book, 1973) is designed to minimize the effect of numerical errors and to retain the correct physics while resolving steep gradients at detonation and shock fronts, flame fronts, and multiphase flow interfaces. It approximates high Reynolds number solutions of the Navier-Stokes equations and treats multiple chemical species in a single velocity field. FAST3D also contains a simple hydrogen-oxygen induction parameter (reduced chemistry) model. The LLNL CHEETAH thermochemical code (Fried and Souers, 1994) predicts temperatures, pressures, detonation products, and energetics generated during explosive detonations by calculating the chemical equilibrium of the reacting explosive products while tracking the thermodynamic state of the expanding gases. We are working towards integration of thermochemical and reactive flow codes while using FAST3D to design a new detonation facility.

3. Modeling A Detonation Facility

FAST3D is presently being used to design an open detonation and open burn facility (ODOBi) at DPG (Fig. 1A). This facility is being designed to handle larger scale (up to 300 lbs) detonations and to serve as a prototype for an operational demil facility. ODOBi will hold the fireball in contact with the plume for a longer time than occurs in an open detonation, thereby increasing byproduct consumption. As an example of design simulation applications, we are using FAST3D to evaluate the overpressures experienced in two octagonal chambers. The first chamber (ODOBi1) has a 2-m opening and a maximum diameter of 8.8-m. The second chamber (ODOBi2), with a 3-m opening and a 9.4-m maximum diameter, has twice the volume of the first chamber. Figs. 1B-1E illustrate pressure contours calculated for a vertical cross-section through each chamber after detonation of 100-lb or 300-lb (TNT equivalent) charges. Peaks in the pressure time series (normalized by ambient pressure) through the center of the top opening presented in Figs. 1F-1G show the effects of shock reflection from chamber walls. Thus, numerical simulation offers the major advantage of design optimization with a minimum of time-consuming testing.

4. Source Characterization Modeling

As the ODOBi design nears completion, we will focus on source characterization. This includes linking reactive flow with thermochemical codes, converting emission factors to dispersion model source terms, and developing a simpler, faster code for real-time operational use. The first step will be the simplification and repackaging of FAST3D to form a Detailed OD Source Characterization (DOD-SC) code. This code will be further simplified to a one-dimensional Workstation Onsite Open-air Detonation (WOOD) code. These codes will form a triad of descending complexity, each depending on the higher-level version and available experimental results for its calibration and validation.

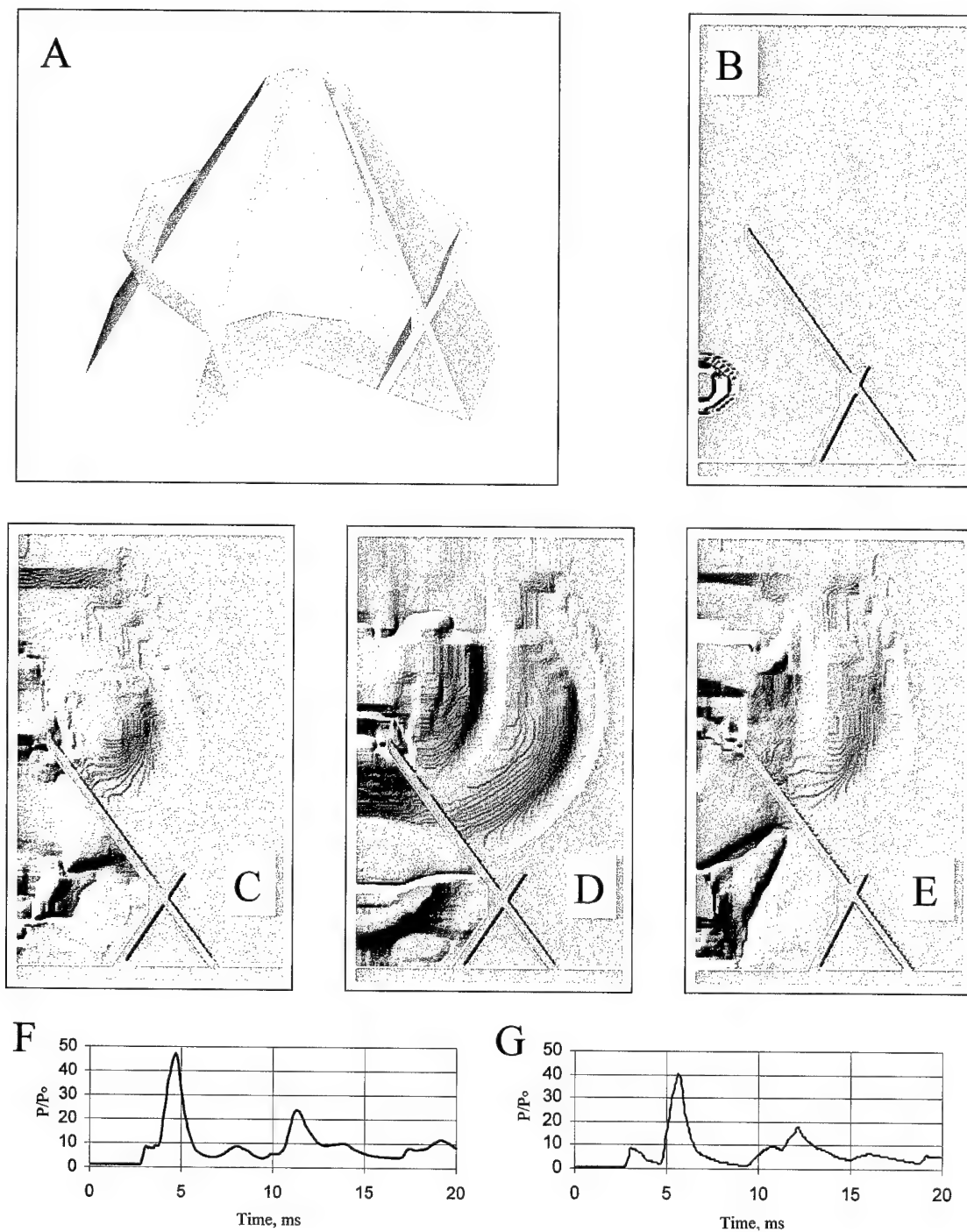


Figure 1. (A) Three dimensional image of ODOBi; (B) cross-section of detonation initiation; (C) pressure contours for a 100-lb charge at 16 milliseconds in ODOBi1; (D) pressure contours for a 300-lb charge at 16 milliseconds in ODOBi1; (E) pressure contours for a 300-lb charge at 16 milliseconds in ODOBi2; (F) overpressure ratio at the rim of ODOBi1 for a 300-lb charge; and (G) overpressure ratio at the rim of ODOBi2 for a 300-lb charge.

DOD-SC and WOOD will be developed using the same general program structure to optimize inter-code information exchange and to allow straightforward upgrades. In addition to providing source characterization for field experiments, proposed OD sites, and facility design, DOD-SC will be used to calibrate a simple, arbitrary topology interconnected network (ATIN) of 1D integration pathways for WOOD. This calibration, which will have to be done for each demil material and OB/OD site, will ensure that WOOD provides sufficiently accurate OD source characterizations while running in just a few minutes on a workstation. Thus, the onsite operator will need only to specify the explosive charge and the source geometry to use WOOD for pollutant emissions predictions. The results will be handed off to nested noise propagation and atmospheric dispersion models. A PC-based OB/OD dispersion model (Bjorklund et al., 1996) is presently available while research continues on a next generation dispersion model.

5. Conclusions

Reactive flow CFD has a multi-faceted role to play in the plan to eliminate dangerous stockpiles of unwanted munitions. This role includes the design of disposal facilities such as ODOBi, evaluation of mitigation strategies, and definition of dispersion model source terms. Our multi-agency measurement and modeling program exemplifies inter-agency partnering in pursuit of a common goal: a safe and secure environment for our Nation's future.

Acknowledgements

The authors wish to acknowledge Susan Gross and Robert Feldman of the DPG Meteorology & Modeling Division for word processing and graphics support, and the helpful comments from Joseph Knox of HQ, U.S. Army Test & Evaluation Command.

References

1. Johnson, M., *Development of Methodology and Techniques for Identifying and Quantifying Products from Open Burning and Open Detonation Thermal Treatment Methods - BangBox Test Series, Volume 1 (Test Summary)*, U.S. Army, AMMCOM, Rock Island, IL 61299-6000, January 1992.
2. Oran, E.S. and Boris, J.P., *Numerical Simulation of Reactive Flow*, Elsevier Science Publishing Co., New York, NY, 1993.
3. Boris, J.P., Landsberg, A.M., Oran, E.S., and Gardner, J.H., *LCPFCT-A Flux-Corrected Transport Algorithm for Solving Generalized Continuity Equations*, NRL/MR/6410-93-7192, Naval Research Laboratory, Washington, DC, April 1993.
4. Boris, J.P., and Book, D. L., "Flux-Corrected Transport I: SHASTA-A Fluid Transport Algorithm That Works," *J. Comp. Phys.*, Vol. 11, pp.38-69, 1973.
5. Fried, L., and Souers, P.C., *CHEETAH: A Next Generation Thermochemical Code*, UCRL-ID-117240, Lawrence Livermore National Laboratories, Livermore, CA, 1994.
6. Bjorklund, J.R., Bowers, J.F., Dodd, G.C., and White, J.M., *Open Burn/Open Detonation Dispersion Model (OBODM) User's Guide. Volume I. User's Instructions; Volume II. Technical Description*, DPG-TR-96-008, U.S. Army Dugway Proving Ground, Dugway, UT, January 1996.

Electronics

STRAIN ENGINEERED SEMICONDUCTOR HETEROSTRUCTURES FOR NOVEL OPTOELECTRONIC DEVICES

Paul H. Shen, Jagadeesh Pamulapati, Michael Wraback, Weimin Zhou,
Monica Taysing-Lara, and Mitra Dutta
U. S. Army Research Laboratory, Physical Science Directorate, AMSRL-PS-PB
Fort Monmouth, New Jersey 07703-5601

Abstract

In this paper we show that in addition to improving the performance of existing components, strain may be used to achieve greater functionality in novel optoelectronic devices. We give as examples three such devices that we have conceived and demonstrated, in the two areas of strain, lattice mismatch induced and thermal expansion coefficient mismatch induced. The higher performance and functionality in these devices demonstrate that strain engineered heterostructures are a very promising area for device research and development.

1. Introduction

The future of the Army is dependent on a highly mobile, seamless, and secure information distribution technology. Any attempt to efficiently transfer dynamic, multifaceted battlefield data requires the implementation of high efficiency, high speed, semiconductor electronic or photonic networks. Strained-layer heterostructures have an impact on nearly all optoelectronic applications and are crucial components in the design of these semiconductor networks. Strain engineering holds the promise of unparalleled advancements in the inchoate domain of battlefield digitization and communication.

Although strained-layer structures have been implemented in major applications such as lasers, photodetectors, and bipolar and field-effect transistors, the strain is mainly utilized to improve the characteristics of an existing device, rather than to offer new applications¹. In this paper, we report several novel kinds of strained structures including, but not limited to, 1) delta strained quantum well structures; 2) variable strain quantum well structures; and 3) anisotropically strained quantum well structures. These novel structures exhibit new electronic and optical properties not observed previously in conventional strained materials, and thus offer new functionalities. The higher performance and functionality in these devices demonstrate that strained heterostructure engineering is a very promising area for device research and development.

2. Lattice mismatch induced biaxial strain

For a strained quantum well of lattice constant a_{QW} grown on a substrate with a lattice constant a_s , the strain is given by

$$\begin{aligned}\epsilon_{xx}(z) &= \epsilon_{yy}(z) = \frac{a_s - a_{QW}(z)}{a_{QW}(z)} \\ \epsilon_{zz}(z) &= -\frac{C_{12}}{C_{11}}[\epsilon_{xx}(z) + \epsilon_{yy}(z)]\end{aligned}\quad (1)$$

In most cases the lattice constant a_{QW} and therefore the strain inside the well do not vary. This constraint can be relaxed by grading of the alloy composition inside the QW and thus changing the $a_{QW}(z)$. For example, if $a_{QW}(z)$ varies linearly in the

quantum well, it is called a variable strained quantum well, while if $a_{\text{QW}}(z)$ has a significant change within a few mono-layers, it is called a delta-strained quantum well.

2.1 Delta-strained Quantum well:

In the delta strained quantum well structure a few monolayers of highly strained material are grown at the center of the quantum well (Fig. 1). The material inserted is chosen such that light hole band is significantly perturbed by the strain while the heavy hole band is not. We have calculated the band structure using the Kohn-Luttinger Hamiltonian including the strain effect. The results indicate that a highly strained layer grown at the center of a lattice matched QW changes the light-hole wavefunction by further decoupling the heavy- and light- hole bands (see Fig. 1) and altering the valence band dispersion (not shown in this paper) such that the heavy-hole effective mass is reduced. A smaller heavy-hole effective mass is desirable in semiconductor laser applications² because it minimizes the carrier density for population inversion while significantly suppressing nonradiative Auger recombination. Although similar results can be achieved with uniformly strained QWs, any improvements are limited due to design compromises between strain and critical layer thickness. The delta-strain approach offers greater design flexibility and enhanced semiconductor laser performance.

The samples used in this study were grown by molecular beam epitaxy on InP substrates. A few monolayers of highly strained (2% compressive) InAlGaAs were inserted in the center of a lattice matched InGaAs/InAlAs 80Å QW.

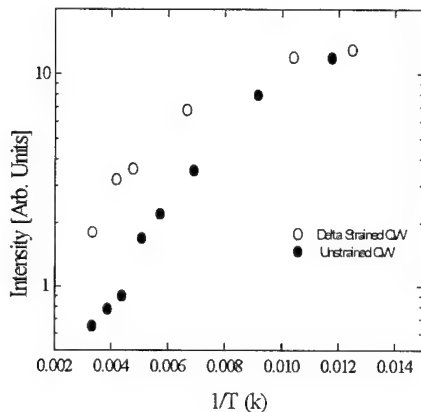


Fig. 3 Integrated PL intensity as a function of temperature. Open circles: delta strained QW; closed circles: unstrained QW

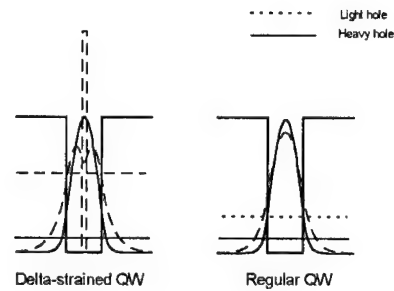


Fig. 1 Illustration of the valence band structure and the lowest heavy- and light-hole wavefunctions of the delta strained quantum well and the regular quantum well.

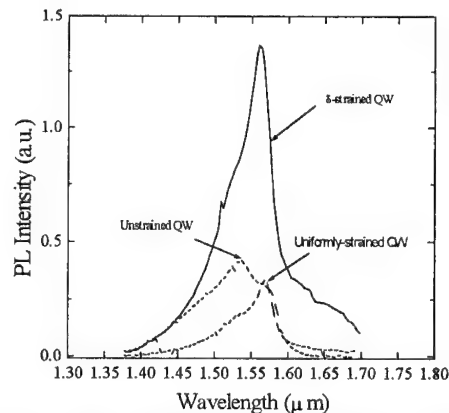


Fig. 2 Room temperature photoluminescence spectra from delta strained QW (solid line), unstrained QW (dashed line) and uniformly strained QW (dotted line).

Two control samples cut from the same wafer and sequentially grown under the same conditions without the delta-strained layer were used as references. One has the same well width as the delta-strained sample, while the other is a 43Å InGaAs/InAlAs QW with a uniform compressive strain of 1%.

Shown in Fig. 2 are the room temperature photoluminescence (PL) spectra from the structures. All the samples have room temperature PL at 1.55μm. However, the integrated room temperature PL intensity from the delta-

strained QW is more than three times stronger than that from the lattice matched QW. We attribute this enhancement of the integrated room temperature PL intensity to the reduction of non-radiative recombination in the delta-strained QW.

Figure 3 shows the temperature dependence of the integrated PL intensity. Although the integrated PL intensity from the delta-strained QW is more than three times stronger than that from the unstrained QW at room temperature, near 77K the magnitudes approach each other. The integrated PL intensity from the delta-strained QW exhibits a weaker (about 3 times) temperature dependence than the unstrained QW, suggesting that the enhancement of the room temperature PL in a delta-strained QW is due to suppression of the non-radiative process rather than enhancement of the radiative contribution.

2.2. Variable-Strained Quantum well

The variable-strain QW structure used in this study consists of a 156Å wide $\text{In}_x\text{Ga}_{1-x}\text{As}$ well, with the InAs mole fraction x graded nearly linearly from $x=0.55$ to 0.35 . This is incorporated into the i(intrinsic)-region of a $\text{In}_{52}\text{Al}_{48}\text{As}$ p-i-n structure (with the p, i, and n layers 5000Å, 7500Å, and 2500Å wide, respectively) nominally lattice matched to InP. Such a linearly varying strain allows the heavy-hole (HH) and light hole (LH) splitting to gradually change from one side of the QW to

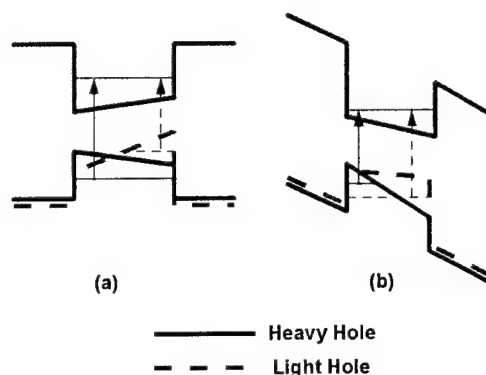


Fig.4. Schematic illustration of the energy bands for a variable-strain quantum well sample; a) without bias, b) with a bias.

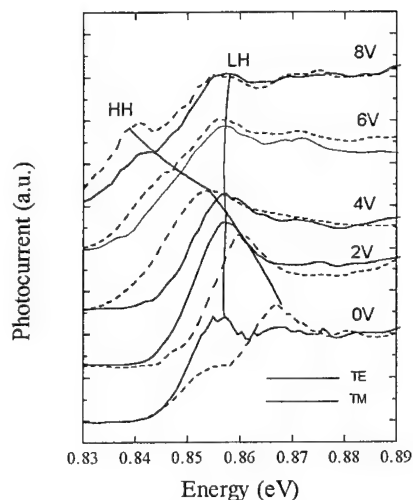


Fig. 5 Photocurrent spectra at different biases, dashed lines: light hole; solid lines: heavy hole

representing the TE mode and the solid lines representing the TM mode. Since the heavy-hole (HH) transition dominates for the TE mode and the light-hole (LH) transition is the only allowed one for the TM mode, the identification of HH and LH transitions is unambiguous. The heavy hole transition displays a strong red shift (30meV) when the bias changes from 0V to 8V, while the light-hole transition exhibits a small blue shift in the bias range 0-3V. Near 3V, the heavy- and light-hole cross each other.

Since the crossing of the 11H and 11L transitions significantly alters the polarization properties of the material, this situation leads to many new opto-electronic device applications such as tunable waveguide polarizers, switches and modulators. Note that

the other (Fig. 4a). Neglecting the strain-induced coupling between the LH and the spin split-off (s.o.) bands, the HH and LH band edges in the QW have opposite potential gradients (opposite fields). When a bias is applied to the QW (Fig. 4b), it reduces the field for the LH band, but increases the field for the HH band. Therefore the first heavy hole band (HH1) will have a red shift, while the first light-hole band (LH1) will have a blue shift due to the quantum confined Stark effect (QCSE). As a result, the transition energy (11L) from LH1 subband to the first electron subband (e1) should increase (blue shift), and the transition energy (11H) from HH1 to e1 should decrease (red shift), making the 11H and 11L cross over possible.³

Shown in Fig. 5 is the low temperature (10K) photocurrent spectra measured at different biases. The measurement was performed in a waveguide configuration, with the dashed lines

in a regular QW the QCSE caused by an electric field can shift HH1 and LH1 bands in the same direction at different rates due to the difference in the effective masses. In contrast, opposing shifts can be realized in a VSQW.

3. Thermally induced anisotropic strain

Although Eq. 1 is valid for strained-layer structures grown on (100) substrates, the lattice mismatch induced strain for any (hhk) orientation is biaxial. Anisotropic in-plane strain ($\epsilon_{xx} \neq \epsilon_{yy}$) can only be achieved if a_s in Eq. 1 has a different value for the x and y directions. In our case it is achieved by bonding a MQW thin film at a temperature T_0 to a host substrate which has a direction-dependent thermal expansion coefficient ($\alpha_x \neq \alpha_y$). At a temperature $T \neq T_0$, a thermally induced strain of

$$\begin{aligned}\epsilon_{xx} &= (\alpha_{x,s} - \alpha_{QW})(T - T_0) \\ \epsilon_{yy} &= (\alpha_{y,s} - \alpha_{QW})(T - T_0)\end{aligned}\quad (2)$$

is induced, where α_{QW} is the thermal expansion coefficients of the quantum well thin film. An in-plane anisotropic strain breaks the rotation symmetry of the valence band at $k_{\parallel}=0$, mixing the heavy and light hole band in the MQW and creating an anisotropic excitonic absorption.⁴ Application of an electric field to these structures results in tunable polarization rotation and phase retardation, characteristics which make these materials suitable for novel device applications. When the field is applied parallel to the growth direction the quantum-confined Stark effect produces a shift of the absorption spectrum⁵, while an electric field applied in the plane of the layers provides an anisotropic quenching of the excitonic absorption.

We have recently fabricated and tested a polarization and phase sensitive modulator. The device is composed of a p-i(MQW)-n structure on (100) GaAs. The i region consists of a 150 period, 80 Å GaAs / 60 Å $\text{Al}_{0.30}\text{Ga}_{0.70}\text{As}$ MQW. A 1000 Å AlAs sacrificial etch stop layer was grown beneath the p-i-n structure. The wafer was first thinned to $\sim 150 \mu\text{m}$. It was then inverted and attached to lithium tantalate (LiTaO_3) at 150°C with a thin layer of UV curable optical adhesive. The sample was then held at 150°C for 24 hours. The GaAs substrate was selectively removed using citric acid and the etch stop layer was removed using a 10% buffered HF etchant. Optical and X-ray measurements have shown that this sample was under an anisotropic strain of $\sim 0.15\%$ at room temperature. The sample was subsequently fabricated into an array using standard photolithographic techniques. Finally, an identical piece of LiTaO_3 was attached to the back of the device substrate, in an orthogonal direction, to compensate for the optical birefringence in LiTaO_3 .

The device was tested with normal incident laser light linearly polarized at 45° with respect to the strain axis. To obtain the maximum tunable phase retardation as well as polarization rotation, the operating wavelength (λ) was chosen at 845nm, which is slightly below the heavy-hole exciton peak at a bias of 14V (off state). The maximum tunable polarization rotation and phase retardation achieved in these experiments were 15° and 37° , respectively. Shown in Fig. 6 is the emergent light intensity at 845nm as a function of applied bias. The open circles are the transmitted intensity $I_o(V)$, which exhibits the simple amplitude modulation of a QCSE device. To enhance the contrast ratio a polarizer with an orientation perpendicular to the polarization of the transmitted beam in the off state (14V) was inserted in the beam path, significantly reducing the transmitted light. By decreasing the voltage towards the on-state (0V), the polarization rotation and phase retardation result in an increase in transmission through the polarizer. Shown in Fig. 6 by the solid circles is the transmitted light after the polarizer as a function of bias. A very high contrast ratio of 5000:1 was obtained due to the fact that the intensity of the light in the off-state was greatly attenuated.

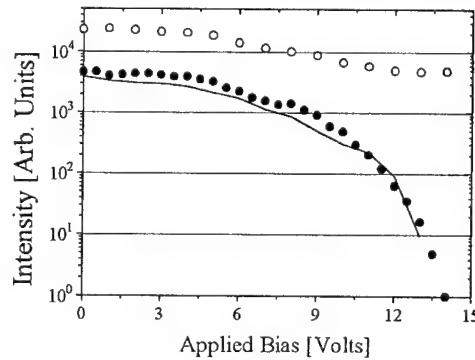


Fig. 6 Transmitted light intensity as a function of applied bias with (closed circles) and without (open circles) insertion of a polarizer.

Finally, we have examined the manufacturability aspects of the ASQW modulator arrays and compared them with the more commonly used asymmetric Fabry-Perot (ASFP) modulator arrays. In the ASQW modulator, the transmitted light intensity $I(V)$ with the polarizer oriented perpendicular to the transmitted beam is given by:

$$I(V) = I_o(V) \{ \sin^2 [\theta(V) - \theta_{off}] + \cos [2\theta(V)] \cos [2\theta_{off}] \sin^2 [\frac{\Delta\phi(V)}{2}] \} \quad (3)$$

where θ_{off} is the polarization rotation angle in the off state, $\theta = \arctan[\exp(-\Delta\alpha d/2)]$, $\Delta\phi = 2\pi\Delta n d/\lambda$, and d is the thickness of the quantum well. The calculated $I(V)$, represented in Fig. 6 by a line, is in good agreement with the experimental results. In an ASFP reflection modulator, the total reflectivity R_T is related to the front mirror reflectivity R_f , the back mirror reflectivity R_b , the round trip cavity absorption $e^{-2\alpha d}$, and the phase change of the optical beam for a single pass through the cavity $\phi = 2\pi n d/\lambda$. It is given by⁶

$$R_T(V) = \frac{[\sqrt{R_f} - \sqrt{R_b} e^{-\alpha d}]^2 + 4\sqrt{R_f R_b} e^{-\alpha d} \sin^2(\phi)}{[1 - \sqrt{R_f R_b} e^{-\alpha d}]^2 + 4\sqrt{R_f R_b} e^{-\alpha d} \sin^2(\phi)} \quad (4)$$

Although in principle the contrast ratio of an ASFP modulator can be as high as an ASQW modulator, the contrast ratio of a ASFP spatial light modulator array is limited because the off-state is extremely sensitive to the total thickness of the QW. For example, in a typical ASFP that has a $\phi = 8\pi$, a 1% thickness variation across the array leads to a $\delta\phi$ which varies by 14° . According to Eq. 4, this corresponds to a maximum contrast ratio of 16. Conversely, the anisotropically strained QW modulator array is less sensitive to any non-uniformities. To achieve maximum contrast in an anisotropically strained QW modulator array, the transmitted light should be linearly polarized ($\Delta\phi=0$) along $\theta_{off} + \pi/2$. Consequently, Eq. 3 implies that one may still obtain a maximum contrast ratio of 3000:1 while assuming the same 1% thickness variation along with an additional 1% variation in the strain.

4. CONCLUSION

We have presented three unconventional strained semiconductor heterostructures: one possessing delta strain, one possessing strain varying along the growth direction, and finally one possessing a thermally induced in-plane anisotropic strain. We have demonstrated that valence band engineering using unconventional strain is very powerful and results in many novel and desirable electronic and optical properties not seen in conventional strained materials. The higher performance and functionality in these devices demonstrates that strain engineered heterostructures are an extremely promising area for device research and development.

REFERENCES

1. Osbourn, G. C., "Strained-layer superlattices from lattice mismatched materials," *J. Appl. Phys.*, Vol. 53, pp. 1586-1589, 1982.
2. O'Reilly, E. P. and Adams, Alfred R., "Band structure engineering in strained semiconductor lasers," *IEEE J. of Quantum Electronics*, Vol. 30, pp. 366-377, 1994.
3. Zhou, Weimin, Shen, H., Pamulapati, J., Cooke P., and Dutta, M., "Heavy- and light-hole band crossing in a variable-strain quantum-well heterostructure," *Phys. Rev. B*, Vol. 51, pp. 5461-5464, 1995.
4. Shen, H., Wraback, M., Pamulapati, J., Newman, P.G., Dutta, M., Lu, Y. and Kuo, H. C., "Optical anisotropy in GaAs/AlGaAs multiple quantum wells under thermally induced uniaxial strain," *Phys. Rev. B*, Vol. 47, pp. 13933-13936, 1993.
5. Shen, H., Pamulapati, J., Wraback, M., Dutta, M., Kuo H. C. and Lu, Y., "High contrast optical modulator based on electrically tunable polarization rotation and phase retardation in uniaxially strained (100) multiple quantum wells," *IEEE Photonics Technology Letters*, Vol. 6, pp. 700-702, 1994.
6. Hu, Kezhong, Chen, Li, Madhukar, Anupam, Chen, Ping, Kyriakakis, Chris, Karim, Zaheed, and Tanguay, Armand R. Jr., "Inverted cavity GaAs/InGaAs asymmetric Fabry-Perot reflection modulator," *Appl. Phys. Lett.*, Vol. 59, pp. 1664-1666, 1991.

NEW LIGHT COUPLING SCHEME AND QUANTUM TRANSITION NOISE OF QUANTUM WELL INFRARED PHOTODETECTORS

K. K. Choi, W. H. Chang
U. S. Army Research Laboratory, PSD
Fort Monmouth, NJ 07724-5601

C. J. Chen, and D. C. Tsui
Dept. of Electrical Engineering,
Princeton University,
Princeton, NJ 08540.

Abstract

Long wavelength infrared detection is critical to the Army's war fighting capability. In this work, we describe a new approach of light coupling to a quantum well infrared photodetector (QWIP) to enhance its sensitivity. With the new approach, the quantum efficiency of a QWIP can be increased by a factor of four, while the dark current is reduced by a factor of two; the overall sensitivity thereby increases by a factor of 5.6. With these improvements, the detector can be operated at a much higher temperature. In addition, the coupling scheme is pixel size and detection wavelength independent, enabling the QWIP used in multicolor detection. In this work, we will also discuss the noise properties of a QWIP, which play a critical role in detector design and optimization.

1. Introduction

Long wavelength infrared detection is an important part of the Army capability. Recently, quantum well infrared photodetector (QWIP) technology has shown great potential in both tactical and strategic arena. High quality 256×256 infrared images in the $9 \mu\text{m}$ range and 128×128 images in the $15 \mu\text{m}$ range have been demonstrated at an operating temperature of 70 K and 45 K respectively. Although these operating temperatures present no obstacles in strategic applications, they are still lower than the tactical requirements. Nevertheless, an improvement in the operating temperature to 80 K will push the technology into system insertion level. Thus, the QWIP technology is in a critical stage, in which a moderate increase in the detector performance will have an enormous impact in the entire defense capability. In this work, we describe a new approach of light coupling, with which the quantum efficiency of a QWIP can be increased while the dark current will be reduced, enabling the detector operated at a higher temperature. In this work, we will also discuss the noise properties of a QWIP, which play a critical role in detector design and optimization. Due to the fact that electron motion in this detector structure is governed by quantum transitions rather than the classical drift-diffusion, the noise gain g_n will be larger than the photoconductive gain g_p , when g_p is small. This finding favors detectors with large g_p , and rules out a whole class of detectors which is under hot pursuit in recently years, namely the p-type QWIPs.

2. Corrugated QWIP

Although the conventional light coupling schemes such as reflective gratings and random reflectors may be useful in large pixel sizes, the measured quantum efficiency in small detectors is only around 10 %. Besides the small coupling efficiency, there are also other technological difficulties. For example, due to the wavelength specific nature of the designs, each detector wafer needs a specific coupler to ensure a perfect match between the detector absorption peak and the wavelength design of the coupler. The uniformity of the grating parameters across the entire wafer also have to be tightly controlled. In the grating design, the grating periodicity has to match the wavelength of light in the medium; this means gratings sensitive in the mid-infrared range or shorter need submicron resolution in grating fabrication, while a long wavelength ($\lambda > 15 \mu\text{m}$) detector pixel can only have very few grating periods in the

coupler. These and other difficulties call for a continuation of light coupling research.

In this work, we will discuss a new coupling scheme. The detector structure is referred to as a corrugated quantum well infrared photodetector (C-QWIP). In this structure, 1-dimensional V-grooves or 2-dimensional crossed grid patterns are chemically etched through the active detector region and down to the bottom contact layer to create a collection of angled facets within a single detector pixel. These facets direct normally incident light into the remaining QWIP material through total internal reflection (TIR) for one polarization and through single-slit diffraction (SSD) for another. The top view and the two orthogonal side views of a single pixel are shown in Fig. 1.

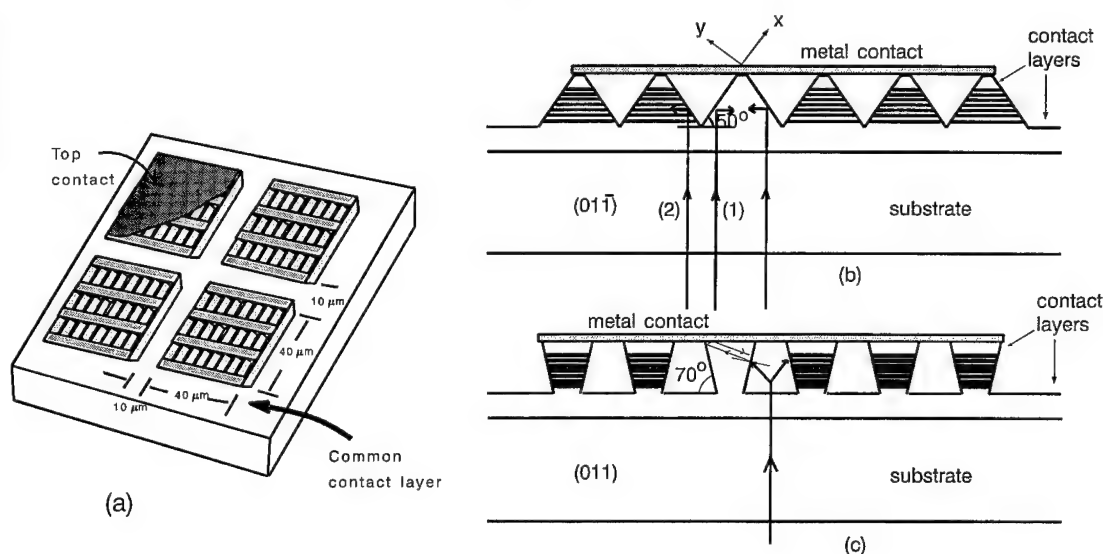


Fig. 1 The 3-dimensional perspective of a C-QWIP array (a) and two side views (b) and (c) of a C-QWIP. For linear patterns, the light is coupled into the QWIP through TIR for wires perpendicular to the $(01\bar{1})$ plane or through SSD for wires perpendicular to the (011) plane. For crossed patterns, the light is either coupled through SSD of the evanescent wave or TIR.

The idea of the C-QWIP is to create a number of sidewalls in a single detector pixel. Under this light coupling scheme, the transverse electric TE mode and the transverse magnetic TM mode coupled into the QWIP through different physical mechanisms. Let us first consider the structure shown in Fig. 1(b), with which only the TM mode is coupled. In this structure, the light is reflected by the slanted sidewalls and makes four optical passes when the substrate is thinned. The ray diagram is shown in Fig. 2. Therefore, for TM polarized light with a thinned

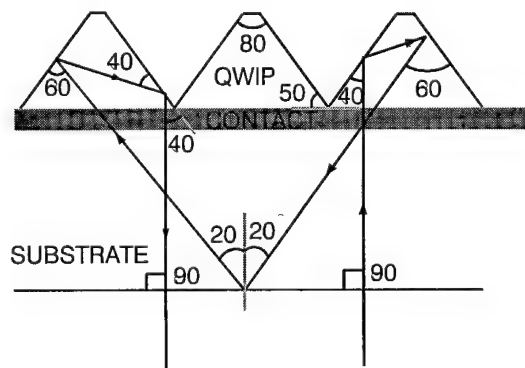


Fig. 2. The ray diagram inside a C-QWIP with 1-dimensional V-grooves. Up to 4 passes can be achieved for a thinned substrate. For an unthinned substrate, we expect the number of passes is 2.

substrate, the intensity inside the QWIP is intensified by 4 times. However, half of the material is removed in this detector geometry, the external quantum efficiency η_{ext} is expected to be 21 %.

For a C-QWIP with mesa grooves illustrated in Fig. 1(c), the light enters the QWIP through single-slit diffraction effect and also through light reflection back from the top metal contact. After undergoing a refraction at the sidewall, the light travels nearly parallel to the layers similar to that of the triangular wires, and we expect the same η_{ext} as in the orthogonal direction. For C-QWIP with 2-dimensional crossed patterns, η_{ext} is expected to be 42 %, four times that of the conventional approaches.

The advantages of a C-QWIP structure can be summarized as follows: (1) The basic coupling mechanisms are size independent. Both total internal reflection (TIR) and single-slit diffraction (SSD) are independent on the number of wires in a pixel. (2) The coupling scheme is more efficient. TIR directs all of the light into a single preferred direction, whereas grating diffraction becomes more isotropic with smaller pixel size. (3) The pathlength is less dependent on the pixel size. In a C-QWIP structure, the light travels nearly parallel to the layers only in the active region and in a much steeper angle in the substrate region. The light reflected from the substrate is more likely directed back into the same pixel. (4) The pixel crosstalk is less as the light is more confined within the same pixel, and the normally incident light will result in a normally exiting light as illustrated in Fig. 2. (5) The C-QWIP is suitable for broad band detection or multi-color detection. It does not rely on destructive interference to eliminate back-scattering of light as in other approaches. And since TIR is wavelength independent, there is significantly less spectral narrowing effect. (6) The dark current is reduced since nearly half of the detector active region is removed. (7) Finally, C-QWIP device processing is simpler. There is no stringent requirement on grating parameters and no thick grating layer to grow. Also, C-QWIP requires only wet chemical etching of the detector material, material damage associated with plasma etching can be avoided.

In the following, the coupling efficiency of a single-color C-QWIP and a two-color C-QWIP will be presented. In these initial detector structures, the substrates of these samples are not thinned, therefore, we expect η_{ext} to be 21 % for crossed patterns. For the single color detector, we measured a C-QWIP with linear V-grooves, a C-QWIP with a crossed pattern, a QWIP with a 45° edge and a QWIP with a plain mesa. Fig. 3 shows the photoresponse from these four detectors. From this measurement, η_{ext} of the detector with a crossed pattern is 15 %, close to the expected value. After substrate thinning and with further device optimization, the theoretical value can be achieved. In the spectral measurement, no significant spectral width narrowing effect is observed in both 1-D and 2-D structures, demonstrating the broadband detection capability of a C-QWIP.

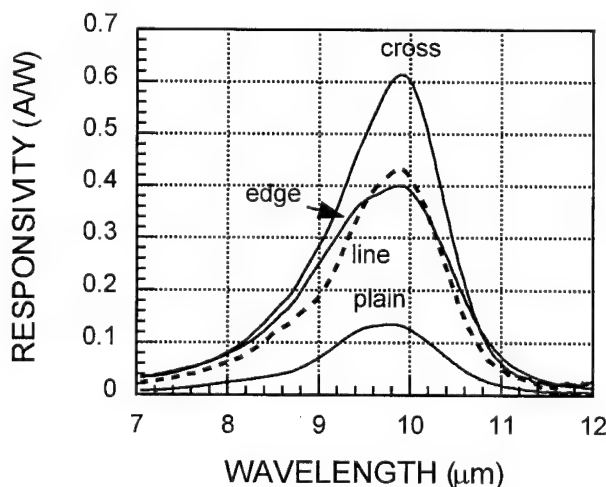


Fig. 3 The responsivity of single-color C-QWIPs with 1-D and 2-D patterns in comparison with that of an edge coupled detector and a detector with a plain mesa.

The responsivity R of the two-color detectors is shown in Fig. 4. A common R ratio of 1.2 is obtained between detectors with 1-D grooves and with edge coupling in both detection windows, manifesting the wavelength independence of a C-QWIP. The present result demonstrates the capability of a C-QWIP in multi-color detection. At the same time, spectral width narrowing effect is very small in both windows. This optical coupling improvement, together with the dark current reduction, leads to an increase of detector performance.

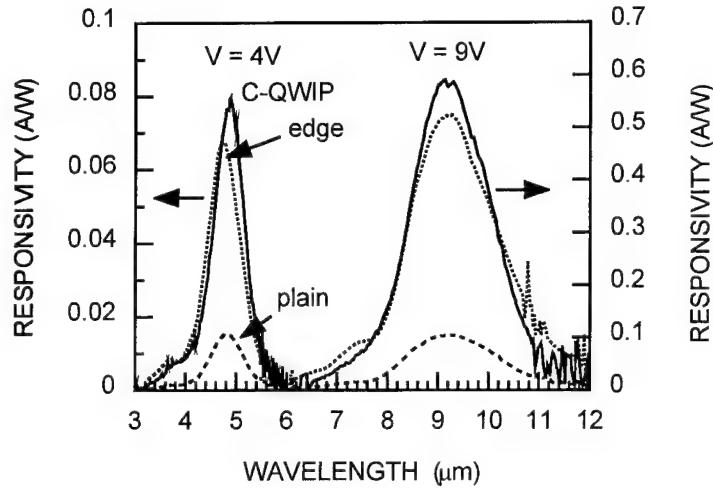


Fig. 4 The responsivity of a two-color C-QWIP in comparison with that of an edge coupled detector and a detector with a plain mesa.

3. Reduction of Photoconductive Gain in QWIP

In this work, we show that there is a reduction of photoconductive gain g_p in quantum well infrared photodetectors from its classical value. The reduction is caused by the quantum nature of electron transport in these structures. On the other hand, the generation-recombination noise gain g_n is unaffected by the transport model, and remains to be the same as that of a classical photoconductor. The reduction of g_p leads to an apparent noise increase in these structures, *ie.* g_n deduced from the noise measurement is larger than g_p deduced from the photoconductivity measurements. We compared the present theory with existing experimental data, and found reasonable agreement.

After a photoexcitation, the dwell time t_d of a photoelectron in the original well depends on the physical picture of electron transport in a QWIP. If an electron transfers from a well region to the adjacent barrier region through quantum transitions, t_d will be the same as that of the subsequent wells, denoted by t_w . On the other hand, if the electron motion is governed by classical drift-diffusion, and if on average electrons are created at the center of the well, $t_d = t_w/2$, half of that of the subsequent wells. Other than these two cases, we also consider a direct case, in which the electron is excited directly into the barrier region, so that $t_d = 0$. In Table I, we list the theoretical value of g_p in terms of the single well capture probability p_c in different cases.

	(a) quantum	(b) direct	(c) classical
g_p	$\frac{1}{N} \frac{1-p_c}{p_c}$	$\frac{1}{N} \frac{1}{p_c}$	$\frac{1}{N} \frac{\sqrt{1-p_c}}{p_c}$

Table I. The value of g_p in different model, N is the number of wells in a QWIP.

On the other hand, g_n is identical in all of the cases, independent on the dwell time of the first well. It is given by

$$g_n = \frac{1}{N} \frac{1 - \frac{p_c}{2}}{p_c} . \quad (1)$$

In fact, the identical g_n is not accidental, but follows a general principle that if a random variable g is increased by a constant value c , the average value will be increased from $\langle g \rangle$ to $\langle g \rangle + c$, but the variance $\langle [(g+c) - (\langle g \rangle + c)]^2 \rangle = \langle (g - \langle g \rangle)^2 \rangle$ should remain the same. Fig. 5 shows the theoretical ratio $K \equiv g_n/g_p$ as a function of g_n for the three different cases.

In order to distinguish these three possible models, the experimental data from Xing *et al.*¹ were analyzed. The data contain both g_n and g_p , from which the experimental K can be deduced, and is plotted in Fig. 5 for two bias polarities. For a large g_n , which occurs at a large applied voltage V , there are little quantitative differences among different models as K approaches 1 in all cases. But for a small g_n occurring at a low V , the differences become large. In this low V regime, the direct model (b) and the classical model (c) predict either a reduction of K or almost no change in K , which are inconsistent with the experimental increasing K as V decreases. Therefore, the quantum model of curve (a) offers the most plausible explanation of the data.

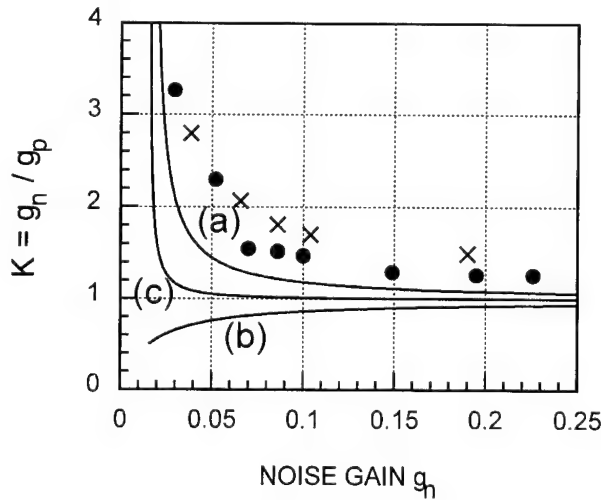


Fig. 5 The theoretical K in the quantum model (a), in the direct model (b), and in the classical model (c) as a function of g_n . The figure also shows the experimental K for the detector with positive (circles) and negative (crosses) voltages.

In conclusion, the C-QWIP structure offers a number of advantages over other coupling schemes, and in many ways, it behaves as a detector having normal incident absorption. Therefore, the polarization dependence of intersubband absorption has been overcome. We also found that the value of g is dependent on the transport model. By comparing with the existing experimental data, the quantum model is shown to be most plausible. The increase of K has a negative impact on QWIPs with a small g , such as p-type QWIPs.

References

1. B. Xing, H. C. Liu, P. H. Wilson, M. Buchanan, Z. R. Wasilewski, and J. G. Simmons, *J. Appl. Phys.* **76**, 1889 (1994).

Advances in Quasi-Optical Power Combiners Provide Path to Radar and Communications Enhancements

J. F. Harvey*
US Army Research Office
Research Triangle Park, N. C. 27709

M. B. Steer, J.W. Mink, H.-S. Hwang, and T. W. Nuteson
North Carolina State University
Raleigh, N. C. 27695

A. C. Paolella
US Army Communications-Electronics Command
Ft. Monmouth, N. J. 07703

Abstract

Recent advances in the state-of-the-art for quasi-optical power combining systems are described, with significant potential impact on military electronics systems. A novel Green's function, which efficiently describes the electromagnetic propagation in a complete 3 dimensional quasi-optical system, was developed and applied in a Method of Moments computer algorithm to form the analytical engine of a powerful Computer Aided Design system. In this method, the overall Green's function describing the propagation through the quasi-optical sequence of lenses and amplifier or oscillator arrays is decomposed into two parts, describing the quasi-optical paraxial fields and the remaining radiated fields respectively. The Green's function was computed in part in the spectral domain to avoid singularities and in part in the spatial domain to efficiently calculate the effects of components well separated in space. Accurate computational results were obtained for the first time for a finite metal grid array in a lens system. These results are verified by experiment. In addition, a highly innovative 2 dimensional planar combining architecture was demonstrated for the first time, with significant potential advantages in size, heat dissipation, and ruggedness over conventional quasi-optical architectures which use large 3 dimensional spatial volumes for the power combining. Efficient combining of a four element array of oscillators and 14 dB gain in a four element amplifier array were demonstrated in initial experiments at X-band.

1. Introduction

Recent advances in the simulation and modeling of quasi-optical (QO) circuits and in the entirely new technology of 2 dimensional QO power combining promise significant progress in bringing the advantages of solid state monolithic technology to moderate and high power millimeter wave applications as envisioned to satisfy Force XXI requirements for radar, communications, and electronic warfare. As a result of its potential to provide a solid state semiconductor source of high power at millimeter wave frequencies, QO power combining plays a major role in the new DARPA MAFET Program for military applications of microwave and millimeter wave technology. The first full-wave computer model capable of simulating far-field and near-field propagation patterns and of accurately determining the propagation effects and driving point impedances in a finite QO grid or array has been developed. In addition, two-dimensional QO power combining of oscillators and amplifiers has been demonstrated for the first time in a dielectric substrate, as has a complete 2D amplifying system. These two advances are a major step toward the replacement of bulky, expensive traveling wave tubes and inefficient antenna array circuits with small, cheap, and reliable integrated millimeter-wave QO circuits. This 2D QO power combining technique results in even lighter, more compact, more rugged, and potentially more efficient systems than can conventional free-space QO systems currently under development.

QO techniques provide a means for combining power from numerous solid-state millimeter-wave sources without using lossy metallic interconnections. Power from the sources in an array is combined over a distance of many wavelengths to channel power predominately into a single paraxial mode, see figure 1. A major characteristic of the conventional QO circuits is the use of 3D free space for the power combining. We have built on this concept to develop a QO system which combines power in two dimensions. Our results demonstrate for the first time 2D power combining in a planar dielectric substrate using amplifier and oscillator arrays. Furthermore, our experiments indicate that the 2D geometry can lead to circuits which are less sensitive to instabilities and which are easier to optimize for efficiency as well as being smaller and lighter. Recent advances in conventional QO arrays have demonstrated circuits near X-band with power added efficiencies as high as 85% and with output powers greater than 10 W. A monolithic array has demonstrated 3 dB gain at 40 GHz. However, progress toward large, high-powered, efficient 2D or 3D arrays is hampered by the relatively crude state of design technology including the lack of suitable computer aided engineering tools. In particular, the many unit active circuits in a large array cannot be individually optimized for efficiency and stability. This is because no simulation process has been developed to model impedances and arrive at stability criteria for a finite array where most of the array elements see different circuit conditions.

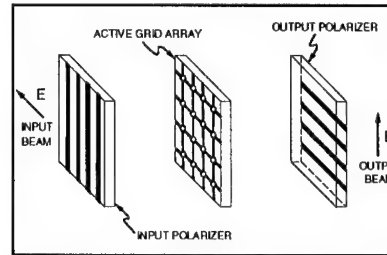


Figure 1. 3D grid array quasi-optical power combining amplifier with input and output polarizers.

The essential component of QO modeling is development of circuit-level models of QO structures. The conventional modeling approach using moment method techniques requires the derivation of a potential Green's function. In this paper a full-wave moment method implementation is developed for the analysis of QO systems. A series of developments culminate in a straight forward methodology for developing a novel Green's function of a QO structure. The electric field dyadic Green's function of a QO system is derived in two parts: one part describing the effect of the QO paraxial fields and the other part describing the remaining radiated fields. This form of the dyadic Green's function is particularly convenient for QO systems because of its relative ease of development. It did, however, necessitate the development of an advanced Method of Moments (MoM) approach combining spatial domain and spectral domain techniques. With this formulation the field solver can be conveniently used in the development of circuit-level models of QO system. Measurements and simulations were performed in free space for a 5×5 QO grid, showing that there is significant mutual coupling between grid elements and that the driving point impedance can vary significantly for different positions on the grid. The ability of our technique to accurately treat these variations and coupling is a major advantage over other modeling techniques attempted for QO structures. The real value of this modeling technique will come from the simulation of much larger arrays.

A very promising alternative to the QO systems which combine power in the three-dimensional volume of free space, as depicted in figure 1, is a two-dimensional combining system such as is illustrated in figure 2. As an inherently 2D planar architecture it is more amenable to photolithographic reproduction than conventional 3D QO structures, and it is ideally suited to the use of MMIC technology. Heat-sinking is easily achieved since the active devices are located at the bottom of the structure and can be placed in thermal contact with a metal ground plane. In contrast, 3D QO systems have the yet unresolved problem of heat dissipation, since the active devices are suspended in the 3D volume. By confining the electromagnetic fields in the 2D structure of the substrate

with a high dielectric constant, the 2D architecture can concentrate the energy into a much smaller volume than the conventional 3D architectures, and thus should provide a smaller and lighter weight system. Further, the 2D structure is essentially a solid, metal backed, slab of material with devices and electromagnetic components locked into fixed relative positions. It thus offers advantages in ruggedness and resistance to component misalignment. As frequencies rise into the mm-wave range, one of the principal loss mechanisms for transmission line or antennas on a substrate is the conversion of the electromagnetic energy into substrate modes. Our 2D combining architecture makes use of the substrate modes, in a controlled fashion, to do the power combining and signal routing. Here the device and antenna structures on the lower surface of the substrate are optimized to radiate into the substrate modes, rather than attempting to avoid or reduce such radiation. For mm-wave frequencies it is more natural to perform signal processing functions in a distributed manner on the electromagnetic fields in the substrate than to attempt to confine the electromagnetic energy in fine metallic structures in order to process it by

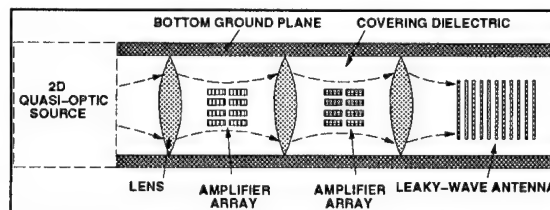


Figure 2. Two dimensional quasi-optical power combining system.

localized circuit elements.

Experimental confirmation has been successfully achieved for the viability of each of the essential components of the QO 2D power combining system shown in figure 2. A 2D oscillator array was demonstrated as a possible source of mm-wave radiation. A 2D amplifier array has been demonstrated with high gain. In a system such an array would boost the power level of the radiation to useful levels for communications, radar, seeker, or electronic warfare applications. Finally, a leaky wave antenna on the surface of the substrate has demonstrated a frequency scanned beam to couple out radiation from the combining substrate either as a scanned antenna array element or to couple into another substrate amplifier combiner as a component in a stacked architecture. It is also possible (although not yet demonstrated) to combine a 2D array of frequency multipliers in a substrate, as an alternative to generating the mm-wave radiation at the fundamental frequency, in order to overcome frequency limitations of solid state transistor technology.

2. Analysis of a Finite Quasi-Optical Grid Structure

The MoM technique is used to solve the integral equation:

$$\mathbf{E}(\mathbf{r}) - \int \mathbf{G}_{\mathbf{E}}(\mathbf{r}, \mathbf{r}') \mathbf{J}_s(\mathbf{r}') d\mathbf{r}' \quad (1)$$

for the quasi-optical system shown in figure 3, where \mathbf{E} is the electric field, $\mathbf{G}_{\mathbf{E}}$ is the electric field dyadic Green's function, \mathbf{J}_s is the electric current density, and \mathbf{r} and \mathbf{r}' are the observation and source coordinates respectively. In this type of electromagnetics problem, \mathbf{J}_s is an unknown current which is generated in response to an incident electric field and which

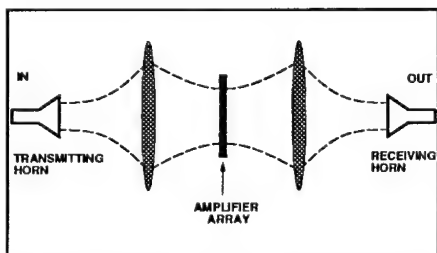


Figure 3. 3D quasi-optical grid lens system.

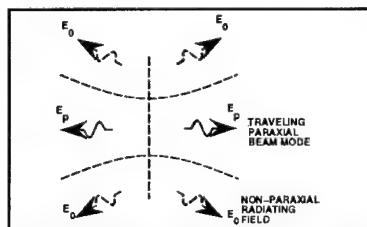


Figure 4. Paraxial and radiating fields excited by a current on the grid.

generates an electric field in turn. The overall electric field, \mathbf{E} , is the superposition of the incident field and the fields generated by currents within the system. The overall field, \mathbf{E} , must satisfy appropriate boundary conditions. Equation (1) is solved by expanding \mathbf{J}_s in basis functions and inverting the resulting matrix equation.

Our technique for deriving the Green's function for such a QO

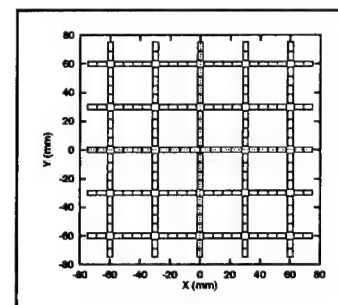


Figure 5. Metal grid for quasi-optical array.

system is based on our previous work on QO open resonator systems. The special insight in our technique is the derivation of $\mathbf{G}_{\mathbf{E}}$ in two parts, describing the free space field, \mathbf{E}_0 , and the paraxial traveling wave QO field, \mathbf{E}_p , see figure 4. There is a significant part of \mathbf{E}_0 which contributes to the paraxial field. When active devices are operated in the array, this part of \mathbf{E}_0 is the production or amplification of the QO field, which occurs at the device locations. This overlap between \mathbf{E}_0 and \mathbf{E}_p must be carefully accounted for in the derivation to avoid double counting a large field component. \mathbf{E}_0 includes radiation and near-field interactions between array elements. The Green's function for \mathbf{E}_0 is evaluated in the spectral domain to avoid the singularity that occurs when \mathbf{r} and \mathbf{r}' are equal. The Green's function for \mathbf{E}_p is more efficiently evaluated in the spatial domain, because it involves more distant interactions. The paraxial field, \mathbf{E}_p , is expanded in Hermite-Gaussian traveling wave modes, which are the natural modes in a paraxial lens system. An advanced MoM approach was formulated which combines the spectral domain and spatial domain techniques. Our new technique is equally

applicable to either conventional 3D QO systems or to 2D systems, with the appropriate representation of the Green's function.

Figure 5 shows the metal grid of the array appearing at the midpoint of the QO system of figure 3. In the actual QO circuit, active devices would be connected at the crossing points of this grid. In order to validate the computational electromagnetic model, measurements of the driving point reflection coefficients from the grid intersections without active

devices, in the lens-array system shown in figure 3, were compared with calculated quantities. The convex lenses were made of Rexolite 1422 ($\epsilon_r = 2.56$) with a diameter of 45.72 cm, focal length of 58.74 cm, and were spaced at twice the focal length.

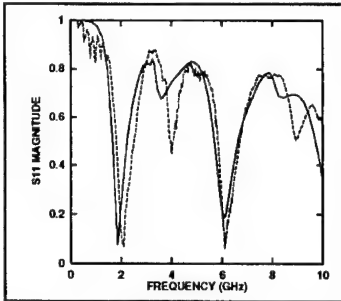


Figure 6. Driving point reflection coefficient magnitude.

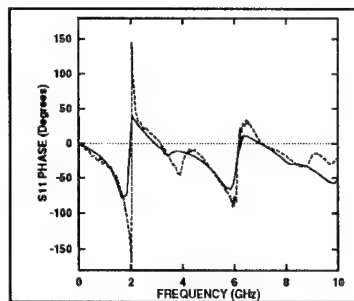


Figure 7. Driving point reflection coefficient phase.

The results shown in figures 6 and 7 demonstrate excellent agreement between calculation (solid lines) and measurement (dotted lines) for the driving point reflection coefficient at the center intersection in the full 5×5 grid of figure 5. Other results show significant differences in the driving point response for a single unit cell vs. a more complete grid. The major differences are due to additional resonances which can be excited in the larger grid structure. Measurement and calculation of the different driving point responses at different grid intersections (eg. center, corner, edge, etc.) show clearly that the various intersections on the grid array are not equivalent and do not see equal driving point

impedances (Nuteson, et al., 1996).

3. Planar Quasi-Optical Power Combining

The components in the QO 2D power combining system shown in figure 2 have all been separately demonstrated at X-band. The 2D oscillator array produced efficient power combining, the 2D amplifier array yielded high gain, the lens system for substrate mode control was successful, and the leaky wave antenna structure proved to be an efficient way to couple the electromagnetic energy out of the substrate with frequency scannable beam control. It should be emphasized that these experiments were designed to demonstrate concept viability and that the circuits and devices were not fully optimized.

Figure 8 shows a four element MESFET oscillator array in a 2D QO resonant cavity. The MESFET was connected directly into the structure of a Vivaldi tapered slot antenna element, which effectively isolated the active device from fields behind the antenna. The resulting oscillator spectra with one to four unit oscillators showed that the power level rises by 3 dB for two oscillators and 6 dB for four. This indicated highly efficient combining of the oscillator output energy, and may point to a major advantage over conventional 3D QO combining of oscillator arrays.

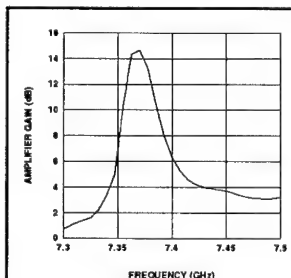


Figure 10. Small signal gain vs frequency (GHz) for 2D quasi-optical amplifier array.

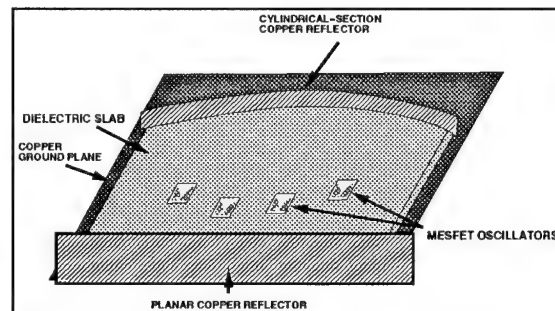


Figure 8. 2D quasi-optical oscillator array.

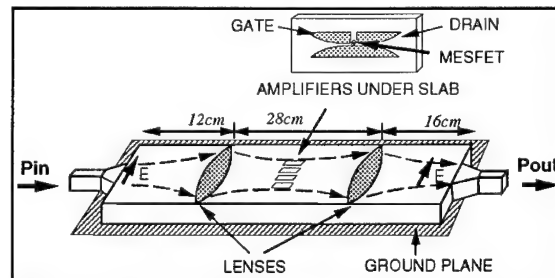


Figure 9. 2D quasi-optical amplifier array.

The four element MESFET amplifier array, shown in figure 9, demonstrated 15 dB small signal gain and roughly 10 dB in saturation (see figures 10 and 11). Each amplifier element consisted of an active device with input and output Vivaldi antennas, which are very

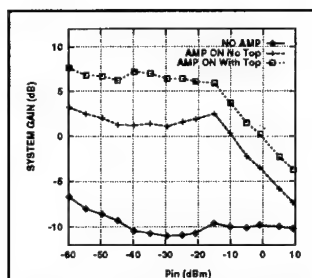


Figure 11. Gain vs input power (dB_m) at 7.35 GHz for 2D quasi-optical amplifier array.

effective in isolating the traveling waves going in opposite directions within the substrate. The MESFET amplifiers were biased at low currents in order to maintain amplifier stability. These active antenna elements were placed under the substrate and above the ground plane, as shown in figure 9, so they would not distort the fields or scatter them from the top surface. The dielectric substrate material was Rexolite ($\epsilon_r = 2.57$) and the lenses were Macor ($\epsilon_r = 5.9$), with focal lengths of 28.54 cm. The substrate width was 30 cm. The substrate thickness was roughly half a wavelength, with the energy propagating in a TE Gaussian beam mode. The use of a TE mode reduces scattering loss because its field value is very small at the surface. The TE mode also is less lossy than the TM mode when a metal cover is placed on top of the substrate, resulting in a 3.5 dB decrease in system loss (see figure 11). Other sources of system loss have been identified, and these appear to be amenable to significant reduction. An additional 4 dB increase in gain can be realized by using concave lenses of air rather than convex lenses of Macor, because the concave shape reduces scattering at the lens. An additional 2 to 4 dB can be recovered by reducing scattering elsewhere in the system.

(Hwang, et al., 1996). In addition, the leaky wave antenna structure was tested (in a non-optimized architecture) as a method to couple the radiation out of the substrate. The output coupling of the energy was efficient, with a 15° scan angle for a 1 GHz (10%) frequency change.

4. Conclusions and Future Direction

A full-wave method of analysis for QO architectures was developed, based on a novel Green's function and an advanced formulation of the MoM. This approach has major advantages over existing full-wave methods. It accurately and efficiently treats a finite array structure, rather than analyzing a unit cell and assuming infinite periodicity. It separates the fields responsible for near interactions from the fields responsible for longer range interactions, and treats each in the most efficient computational domain. It is broadband (from DC to any frequency). It includes conductor and diffraction losses. It accurately treats 3D and 2D QO structures. This approach is also applicable to complicated nonquasi-optical systems, as well. Experimental verification of the model indicates a highly accurate simulation, and also demonstrates the potential inaccuracy in assuming all unit cells of the array see the same driving point impedance and underscores the requirement to model the full finite array. This modeling technique forms the basis for a powerful CAD system which will significantly enhance the ability of circuit designers to design efficient, high power QO arrays for a variety of military applications. Interaction of this electromagnetic solver with active device simulation is now being addressed.

The components for a complete 2D QO system for generating mm-wave power have been demonstrated at X-band. Efficient oscillator arrays, high gain amplifier arrays, substrate lens propagation control, and output coupling via a leaky wave antenna structure were demonstrated. These arrays are being optimized for high power application and MMIC amplifier chips are being directly inserted in the antenna elements. The generation of high power at X-band and mm-wave frequencies, with an architecture that is smaller, lighter, and more rugged than conventional 3D quasi-optical systems, is being addressed.

In addition to the prospects for high power from solid state integrated circuits, these results present the vision of an entirely new circuit integration architecture. Electromagnetic energy would be processed in a distributed manner confined in dielectric substrates. Rather than fighting the limitations of metallic transmission lines and the problems of unwanted substrate modes, engineers would use the substrate modes to optimize circuit performance and power handling capability. Distributed techniques such as photonic band gap components and traveling wave device structures would become the natural architectures for the millimeter wave regime, and devices and EM fields would have to be treated together on an equivalent footing.

References

1. Nuteson, T.W., Monahan, G.P., Steer, M.B., Naidshadham, K., Mink, J.W., Kojucharoff, K., and Harvey, J., "Full-Wave Analysis of Quasi-Optical Structures," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-44, May 1996.
2. Hwang, H.-S., Nuteson, T.W., Steer, M.B., Mink, J.W., Harvey, J., Paoletta, A., "Two-Dimensional Quasi-Optical Power Combining System Performance and Component Design," *IEEE MTT-S Int. Microwave Symp. Dig.*, June 1996.

Multi-Process Synthesis of Novel Ferroelectric Oxide Ceramic Composites for Use in Phased Array Antennas

L.C. Sengupta*, E. Ngo, S. Sengupta and S. Stowell
U.S. Army Research Laboratory
Materials Directorate
APG, MD 21005-5069

Abstract

Various composites of barium strontium titanium oxide (BSTO) combined with other nonelectrically active oxide ceramics have been formulated for its application in phased array antennas. In general, the composites have adjustable electronic properties which can be tailored for use in phased array antennas and other phase shifting devices. The dielectric constant and the loss tangents have been reduced to enhance the overall impedance matching and thereby lowering the insertion loss of the device. In addition, the overall tunability, the change in the dielectric constant with applied voltage, is maintained at a sufficiently high level. In order to address a broad frequency range in the microwave regime, the composites have been fabricated in bulk ceramic, thick films, and thin film form. This paper will present a complete evaluation of the processing, materials characteristics, and breakthrough performance levels of these novel ceramic composites for use in Army electronic devices.

1. Introduction

A phased array refers to an antenna configuration composed of a large number of elements that emit phased signals to form a radio beam. The radio signal can be electronically steered by the active manipulation of the relative phasing of the individual antenna elements. This electronic beam steering concept applies to both the transmitter and the receiver.

Phased array antennas are advantageous in comparison to their mechanical counterparts with respect to their speed, accuracy, and reliability. Complex tracking exercises can be maneuvered rapidly and accurately with a phased array antenna system. These antennas are currently constructed using ferrite phase shifting elements. However, due to the circuit and power requirements necessary to operate these present day antennas, these are extremely costly, large and heavy. Therefore, the use of such antennas has been limited primarily to military applications where the cost can be justified for such capabilities. In order to make these devices practical for many other commercial and military uses, better materials for antenna elements of the antenna must be developed.

The optimal electronic parameters of the dielectric material of choice for phased array applications should be: (a) moderate to low dielectric constant at microwave frequencies, (b) low electronic loss tangents ($\tan \delta$), and (c) high dielectric tunability (change in dielectric constant with applied voltage). The most difficult property to control has been the loss tangent of the material. Various methods have been investigated to lower the dielectric constant of BSTO.¹ However, none of these methods have substantially lowered the intrinsic loss of the material. This report will review the fabrication and characterization of novel BSTO composites. These composites exhibit low loss tangents at microwave frequency such that they can be realistically ($< 1-2$ dB insertion loss) incorporated into phased array antennas. In addition to the discussion of the electronic and microwave properties of the various novel composites, studies on aging / fatigue of the bulk ceramics will be presented.

The thickness of bulk ceramics (>3 mils) limits the operating frequency to around 15 GHz. To extend the frequency range one must investigate techniques to process thick and thin films of the ferroelectrics. Results of thick film fabrication will be discussed via tape-casting. In order to further increase the operating frequency of these devices, thin films were fabricated by the pulsed laser deposition (PLD) and Metallo-Organic Deposition (MOD) methods. The

thin films were deposited on metallized single crystal substrates. The top electrodes were then deposited by a sputtering process to form the vertical capacitor structure. High tunabilities can be achieved in these thin films with much lower applied voltages than those applied to the bulk.

A rigorous investigation of the material characteristics have been performed on all of the specimens included in this study. The microstructures including grain size and phase analysis have been examined using SEM and x-ray diffraction (glancing angle x-ray diffraction was obtained for the thin film materials). Composition and dopant substitution sites have been investigated using Fourier Transform Raman spectroscopy. The analysis of the phase formation and compositional variations has been related to connectivities and to the electronic properties of the materials. This paper will present a complete evaluation of the processing, materials characteristics, and breakthrough performance levels of these novel ceramic composites for use in Army electronic devices.

2. Materials Fabrication

2.1 Bulk Ceramics

Powder forms of Barium Titanate and Strontium Titanate were obtained from Ferro Corporation, Transelco Division, Pen Yan, N.Y. (product nos. 219-6 and 218 respectively), stoichiometrically mixed to achieve $\text{Ba}_{0.6}\text{Sr}_{0.4}\text{TiO}_3$, $\text{Ba}_{0.55}\text{Sr}_{0.45}\text{TiO}_3$, $\text{Ba}_{0.50}\text{Sr}_{0.50}\text{TiO}_3$, $\text{Ba}_{0.45}\text{Sr}_{0.55}\text{TiO}_3$ and ball-milled in ethanol using 3/16" alumina media for 24 hrs. The resulting BSTO was then air-dried, calcined at 1100°C and mixed with the oxide (oxide III) in the proper weights (0, 1, 5, 10, 20, 30, and 60 wt%) and ball-milled again in a slurry of ethanol using the alumina grinding media for an additional 24 hrs. The resulting composites were then air-dried and subsequently sieved with a U.S. standard sieve series #60 (250 micron) sieve. The specimens were then uniaxially pressed to a pressure of around 10,000 psi and subsequently sintered at temperatures which were determined previously with a dilatometer. The ceramics were screen printed with ink, Ferro Product #T2026, and fired at 760°C .

2.2 Ceramic Tape-Casting

The tapes were cast onto Teflon coated mylar sheets at a doctor blade setting of 20 mils. This resulted in green body thicknesses from 5 to 13 mils. The tapes were then removed from the mylar carriers. Single layer tapes were either fired at temperatures which had been previously determined or were screen-printed with a conductive ink and then fired. Metallization of the tapes was accomplished by screening a guard ring pattern on one side of the green (unfired) tape and a ground plane on the other side. The ink used for this was product #E1162, 40% Au, 20% Pd and 40% Pt (firing temperature of 1350°C). Also fired tapes (10, 20, 40 and 60 wt% Oxide III) were screen printed with product #E3309 and subsequently fired at 850°C .

2.3 Thin Film Fabrication

2.3.1 Pulsed Laser Deposition

The targets chosen for this work were $\text{Ba}_{0.6}\text{Sr}_{0.4}\text{TiO}_3$ (BSTO) and BSTO with 1 wt.% of oxide III based compounds. The PLD of the ferroelectric thin films was accomplished using a krypton-fluoride excimer laser ($\lambda = 248$ nm) with an average pulse energy of 300 mJ. The substrate to target separation was maintained at 60 mm. The oxygen partial pressure in the chamber was 50 mT and the substrate temperature was 700°C . The compositions of the films were confirmed by Glancing Angle X-ray diffraction (GAXRD). The thin films in this study were deposited on single crystal sapphire substrates coated with Ruthenium oxide (RuO_2). RuO_2 was the bottom electrode in the parallel plate capacitor structure created with the BSTO composite thin film as the dielectric and Pt as the top electrode.

2.3.2 MOD Deposition

The chemical precursors for the MOD films were barium acetate, strontium acetate, and titanium isopropoxide and oxide III acetate. The acetates of barium and strontium were dissolved in glacial acetic acid and titanium isopropoxide was added to the solution with constant stirring. The additional oxide III precursor was added to the final BSTO solution in various molar concentrations in order to vary the oxide III content of the final composite. The final

solution was diluted with 2-methoxyethanol in order to adjust the viscosity and surface tension of the solution. Dust and other suspended particles were removed by filtering through a 0.2 μm syringe filter. The films were spun coated onto commercially available platinized silicon substrates. It was found that 2-methoxyethanol had improved the wettability and uniformity of coating on the substrate. Following the deposition, they were kept on a hot plate in air for about 10 minutes to remove solvents and other organics. The as-pyrolized films ($\sim 300^\circ\text{C}$) were found to be amorphous and the post-deposition annealing of the films were carried out at about 725°C for a period of about 45 minutes in a tube furnace in room atmosphere to attain crystallinity.

3. Electronic Measurements

3.1 Bulk composites

The dielectric constants, ϵ' , $\epsilon = \epsilon' + i\epsilon''$, loss tangent, $\tan \delta = \epsilon'' / \epsilon'$, and tunability $= \{\epsilon'(0) - \epsilon'(V_{\text{app}})\} / \{\epsilon'(0)\}$ (with an applied electric field of 2.0 V/ μm) were determined for all composites at a frequency range of 1 KHz- 1 MHz. Capacitance measurements for all materials were taken using an HP4284A LCR meter and the dielectric constants were calculated. The electronic properties of the BSTO-Oxide III composite ceramics are listed in Table I. All of the BSTO / Oxide III composites exhibit reduced dielectric constants and loss tangents. As shown in the table, these properties can be reduced further by lowering the barium content. The dielectric constants of the $\text{Ba}_{0.6}\text{Sr}_{0.4}\text{TiO}_3$, $\text{Ba}_{0.55}\text{Sr}_{0.45}\text{TiO}_3$, $\text{Ba}_{0.50}\text{Sr}_{0.50}\text{TiO}_3$, $\text{Ba}_{0.45}\text{Sr}_{0.55}\text{TiO}_3$ / Oxide III composite specimens are the most different at lower weight percent oxide III, decreasing with lower barium content, as would be expected from the differences in the Curie temperatures (see Table I). The resulting materials become more "paraelectric" and, as shown in Fig. 1, the values of the loss tangents are reduced with decrease in barium content. The loss tangent of the undoped specimen is not shown in the figure.

However, as shown in Fig. 2 the tunability of the materials is also decreased with decreasing barium content. Again, this is in agreement with the position of the Curie temperatures. For example if one compares the Curie temperature undoped of $\text{Ba}_{0.6}\text{Sr}_{0.4}\text{TiO}_3$, 100°C , and $\text{Ba}_{0.45}\text{Sr}_{0.55}\text{TiO}_3$, -45°C , it obvious that the materials with lower barium content have much lower Curie temperatures, i.e., are more "paraelectric" at room temperature and more closely resemble the electrical behavior of a linear capacitor and therefore have less tunability. Also as shown in the figure, it is evident that the tunability of the materials increases linearly with increasing field strength and the application of even higher electric fields is currently being investigated. It is apparent from the electronic data that the $\text{Ba}_{0.6}\text{Sr}_{0.4}\text{TiO}_3$ and $\text{Ba}_{0.55}\text{Sr}_{0.45}\text{TiO}_3$ / Oxide III composites provide the best combination of reduced dielectric constants, loss tangents, and maintenance of usable tunabilities.

3.1.1 Aging and Fatigue in Bulk Ceramics

Electronic aging was accomplished by determining the drift in the dielectric constant of the specimen at zero and elevated voltages and the return of the specimen to the zero value dielectric constant was observed.⁵ The dielectric constants of the materials were measured at zero bias and then again when a field of +1.5 V/ μm was applied. The field was sustained for 30 minutes and a reading was taken at one minute intervals during this time period. The field was then decreased to zero and the dielectric constant was again recorded. The BSTO-Oxide III sample showed a 0.410% before biasing, a 1.304% drift at bias, and a 0.808 % drift when returned to zero.

The fatigue data was accomplished by determining the dielectric constant of the part at zero bias, then again upon the application of +1.5 V/ μm and when the field was decreased to zero again and then the dielectric was recorded at a bias of -1.5 V/ μm and again as field was decreased to zero.⁵ The above consisted of one cycle and the specimens were subjected to at least two and a half to five cycles. The tunability of all of the specimens, after the initial biasing cycle, is constant throughout application of the fields, even though the specimens experience typical capacitor-like logarithmic damping of their dielectric constants. For pre-conditioning of the materials, the BSTO-Oxide III composites require about 2 to 2 1/2 cycles. Since the dielectric constants of the materials were observed to return to their pre-bias values, the voltage conditioning should be done after each start-up period for the antenna or, if the antenna is left without application of electric field. The time after which the voltage application has to be done has not yet been determined.

Table 1. Electronic properties of $\text{Ba}_{0.6}\text{Sr}_{0.4}\text{TiO}_3$, $\text{Ba}_{0.55}\text{Sr}_{0.45}\text{TiO}_3$, $\text{Ba}_{0.50}\text{Sr}_{0.50}\text{TiO}_3$, and $\text{Ba}_{0.45}\text{Sr}_{0.55}\text{TiO}_3$ / Oxide III composite bulk ceramics, measured at 1 KHz

Barium Content	Oxide Content (wt%)	Dielectric Constant	Loss Tangent	Tunability	Curie Temp ($^{\circ}\text{C}$)
Ba = 0.45	0	1280.83	0.01184	15.20	-45
	1	1013.60	0.00065	3.71	-80
	5	896.97	0.00069	4.11	-65
	10	768.06	0.00068	3.90	-75
	20	418.98	0.00064	2.47	-70
	30	351.71	0.00061	4.78	-70
	60	78.81	0.00049	3.67	-95
Ba=0.50	0	1907.99	0.05538	25.55	-25
	1	1130.38	0.00081	6.34	-60
	5	1104.75	0.00078	5.10	-55
	10	928.01	0.00076	5.48	-55
	20	592.20	0.00073	6.44	-55
	30	413.91	0.00070	5.82	-70
	60	77.52	0.00096	3.66	-95
Ba=0.55	0	2771.73	0.03904	33.40	-15
	1	1173.14	0.00094	6.52	-45
	5	1348.80	0.00092	8.63	-40
	10	1114.02	0.00094	8.88	-40
	20	742.29	0.00085	8.77	-40
	30	477.97	0.00075	9.54	-45
	60	94.59	0.00034	6.46	-50
Ba=0.60	0	5160.64	0.00961	56.30	10
	1	2254.40	0.00126	15.41	-30
	5	2065.40	0.00145	15.00	-30
	10	1527.34	0.00162	16.60	-30
	20	1068.43	0.00194	15.80	-35
	30	678.39	0.00159	14.80	-30
	60	116.86	0.00148	9.99	-55

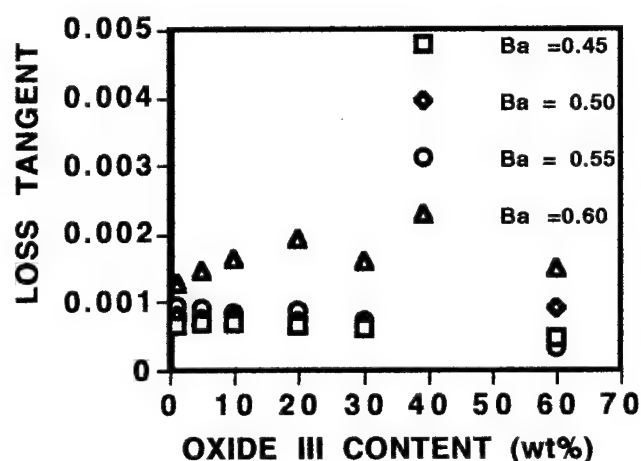


Figure 1. Loss Tangents versus Oxide III Content (wt%) of $\text{Ba}_{0.6}\text{Sr}_{0.4}\text{TiO}_3$, $\text{Ba}_{0.55}\text{Sr}_{0.45}\text{TiO}_3$, $\text{Ba}_{0.50}\text{Sr}_{0.50}\text{TiO}_3$, $\text{Ba}_{0.45}\text{Sr}_{0.55}\text{TiO}_3$ / Oxide III Composites (lines are drawn to aid the eye).

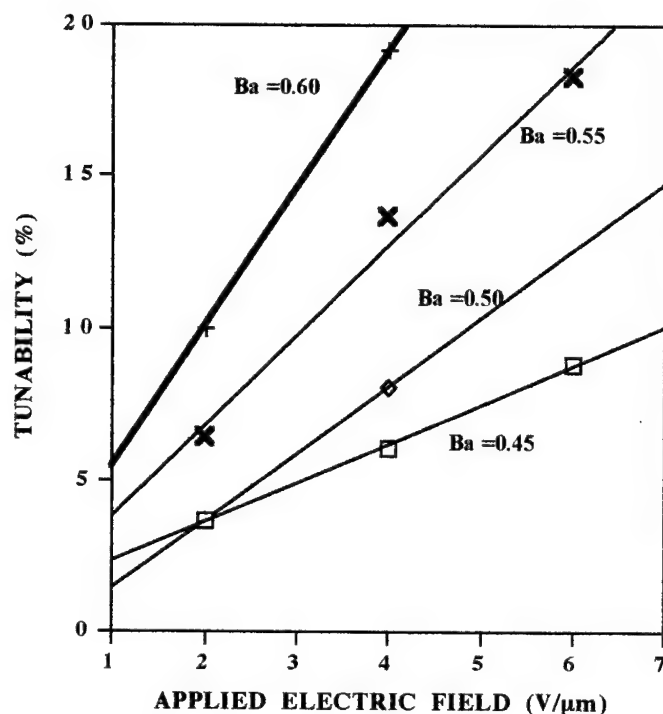


Figure 2. Percent Tunability (2 V/μm) versus Oxide III Content (wt%) of $\text{Ba}_{0.6}\text{Sr}_{0.4}\text{TiO}_3$, $\text{Ba}_{0.55}\text{Sr}_{0.45}\text{TiO}_3$, $\text{Ba}_{0.50}\text{Sr}_{0.50}\text{TiO}_3$, $\text{Ba}_{0.45}\text{Sr}_{0.55}\text{TiO}_3$ / 60 wt% - Oxide III Composites, at (lines are drawn to aid the eye).

3.2 Thick Films

Since higher operating frequencies are required ($> 15\text{GHz}$) thick films of BSTO composites have been fabricated via tape-casting.^{6,7} The electronic properties of the BSTO/ Oxide III tapes are listed in Table 2. The dielectric constants and loss tangents of the tape-cast specimens, similar to the bulk composites, decrease with increase in oxide III content and vary less than 2% with change in frequency (from 1KHz-1MHz). The magnitude of the dielectric constants are very similar to those of the bulk ceramics. The loss tangents of the tapes are also almost the same as those measured for the bulk ceramics. The % tunability of the tapes as well as the bulk material is around 15.00% with 2.00 V/μm at 10 wt% oxide III and only reduces to around 11% at 60 wt% oxide III. This trend was explained previously in the bulk ceramics^{3,4} by the position of the Curie temperatures and the size of the additive (oxide III).

Table 2. Electronic properties of BSTO (Ba = .60) / Oxide III composite single layer tapes measured at 1 KHz

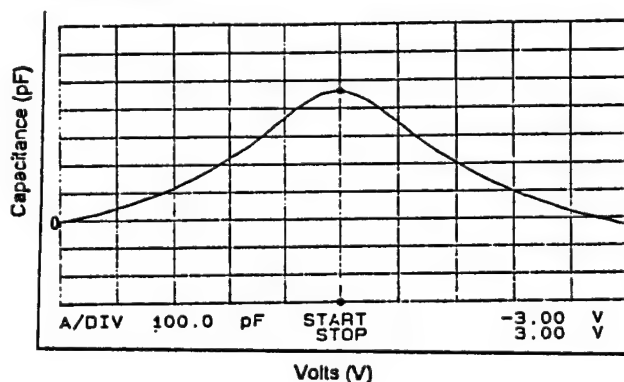
Oxide III Content	Dielectric Constant	Loss Tangent	Tunability (%)	Electric Field
0.0	3192.2	0.0056	43.52	2.00
10.0 (ink 1)	1390.2	0.0015	15.03	2.00
20.0 (ink 1)	616.44	0.0012	15.45	2.00
40.0 (ink 2)	357.30	0.0041	14.00	2.00
60.0 (ink 1)	91.16	0.0008	10.41	2.00

3.3 Thin Films

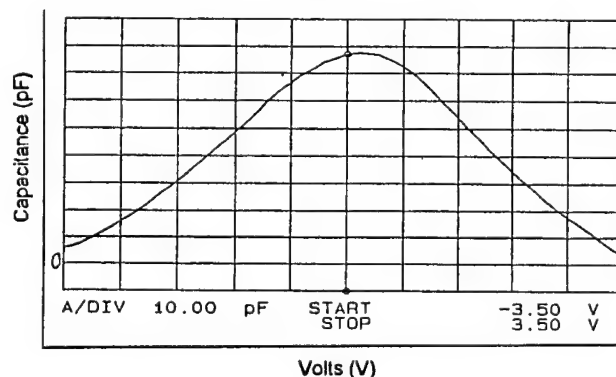
3.3.1 Pulsed Laser Deposited Films

The C-V measurements were performed using a HP 4194 impedance/gain analyzer. All the measurements were performed at 0.5 MHz. The nominal film thickness was 0.6 μm . Fig. 3(a) shows the capacitance versus voltage characteristics at 0.5 MHz for the BSTO (undoped) film deposited on $\text{RuO}_2/\text{sapphire}$. The dielectric constant at zero bias was calculated to be 1200 and the tunability is 23 % at a field of 3 V/ μm . The bulk undoped material has a dielectric constant of 3300 and a tunability of 20 % at 0.73 V/ μm .²⁻⁴ It has been previously shown that the dielectric constant of ferroelectric films are inherently less than the bulk ceramic values due to oxygen defects at the electrode/ film interface.⁸ Also any porosity and/or leakage current in the films will tend to decrease the dielectric constants obtained.

The C-V curve at 0.5 MHz for the BSTO/Oxide III thin film deposited on $\text{RuO}_2/\text{sapphire}$ is shown in Fig. 3(b). The dielectric constant at zero voltage calculated from this curve is 926. The tunability obtained up to 3 V/ μm was 22%. The value for the dielectric constant found in the bulk ceramic target of BSTO/Oxide III was 1276 and a tunability of 16% at 2.32 V/ μm .^{3,4} Therefore, it is obvious that similar trends in the electronic properties as oxide III is added are observed in the PLD films as are observed in the bulk ceramic composites. The tailoring of the electronic properties of BSTO thin films in the low frequency region is possible through the incorporation of metal oxides. The lowering of the dielectric constants, along with a high tunability, plays an important role in the impedance matching of these films into the electronic circuits.



(a)



(b)

Figure 3(a). C-V curve for PLD undoped BSTO thin film 3(b) C-V curve for PLD BSTO /Oxide III thin film.

3.3.2 MOD thin films

The typical small signal dielectric constant and the loss factor for a 0.4 μm thick film processed by the MOD method were 450 and 0.013 respectively. However, with the addition of the oxide III, the dielectric constant of the film decreased to 386 (at 5 mole % of the additive). The dissipation factor was almost reduced by half to 0.007. However, the

tunability of the films were also decreased from 20.7% to 12.7%. It is important to note that the addition of oxide III was found to decrease the leakage current of the capacitor structure from 100 to 10 pA (at 20 mole % additive) at an applied electric field of 100 kV/cm. These results indicate that these BSTO composite thin films are excellent candidates for DRAM applications. A DRAM cell requires a film with high dielectric constant and low leakage current.

4. Fourier Transform Raman Spectroscopy

The Fourier Raman spectra were taken using a Nicolet FT-Raman 950 spectrometer. The system uses a 180° reflective backscattering geometry, a solid state Nd:YVO₄ excitation laser, and an InGaAs detector. An average of 50 -128 scans for each spectra were taken. Since polycrystalline ceramic specimens were used in this study no selective polarization was applied.

The peak position of the higher frequency modes associated with either SrTiO₃ or Oxide III extracted from the FT-Raman spectra for the Ba_{0.45}Sr_{0.55}TiO₃, Ba_{0.50}Sr_{0.50}TiO₃, Ba_{0.55}Sr_{0.45}TiO₃, Ba_{0.6}Sr_{0.4}TiO₃ / Oxide III composite specimens are shown in Fig. 4. As shown in the figure, the peak position is independent of barium content with the addition of oxide III above 1.0 wt%. This indicates that oxide III (> 1 wt%) does not appear to substitute into the BSTO lattice. This conclusion agrees with both SEM/EDS and X-ray powder diffraction results.³ Also, as was shown in Table 1, the position of the Curie temperature for all BSTO / Oxide III composites is approximately unchanged from 1- 30 wt% oxide III additions. Again, this would indicate that the additive is no longer substituting into the BSTO lattice.

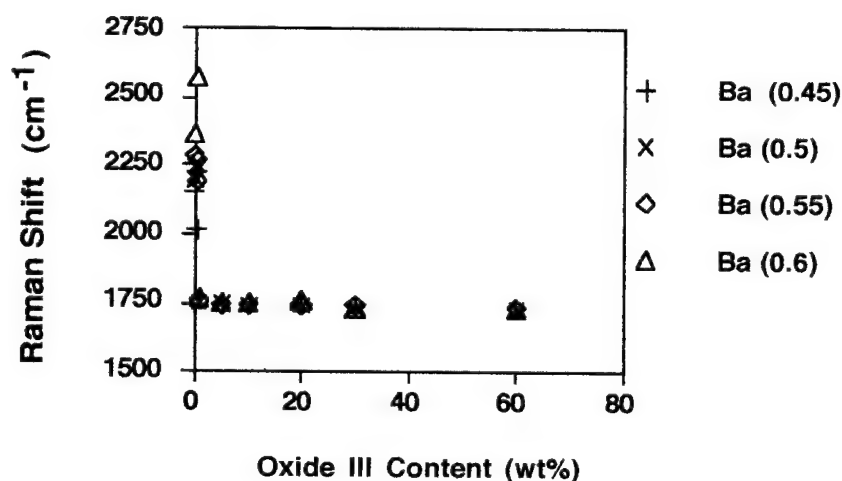


Figure 4. Raman Shift versus Oxide III Content of BSTO / Oxide III Composites for the SrTiO₃ /Oxide III Mode.

5. Microwave Measurements

Microwave measurements were accomplished using a 60 mm X-band cavity, 50 mm X-band cavity, a parallel plate fixture and a resonant mode dielectrometer. These measurements were performed by Dr. R. Geyer, NIST, Boulder, CO. The parallel-plate dielectric resonator fixture was operated using an air gap above the sample to allow for variation in resonant frequency and using rigorous mode-matching techniques for the calculations. The results are shown in Table 3. The uncertainty in the dielectric constant is 0.2% and is related primarily to the uncertainty in the sample dimensions. It is apparent that the low loss features of these composites are maintained in the microwave region. These data represent only three specimens from the BSTO (0.6) / Oxide III composite formulations. The data indicates that the dielectric constants do not change (~5%) from 1 KHz to 10 GHz. As expected for polycrystalline ceramics the loss tangent does increase with frequency. However, the loss tangent is still far lower than any other tunable ferroelectric material (and would represent only ~1 dB of loss in a 10 GHz antenna).

Table 3. Microwave Properties of $\text{Ba}_{0.60}\text{Sr}_{0.40}\text{TiO}_3$ / Oxide III Composites

Oxide III Content (wt%)	Frequency (GHz)	Dielectric Constant	Loss Tangent
30	2.139	646	0.0040
40	1.815	404	0.0042
	3.304	401	0.0051
60	4.581	113	0.0065
	10.02	106	0.012

In addition a four element array with microstrip phase shifters¹¹⁻¹² was designed in collaboration with ARL-PSD using the bulk ceramic BSTO/Oxide III composites (~3 mils thickness). It demonstrated a 360 degree phase shift and less than 3 dB loss at 5 GHz. Coplanar phase shifter elements were also fabricated from the BSTO thin films deposited on sapphire substrates.¹³ The dielectric film was 1 μm thick and the metal was about 0.5 μm thick. The CPW line was 10 μm wide. The structure was subjected to a field of 0 to 20 V in a temperature range of -40C to 180 C at a frequency span of 45 MHz to 35 GHz. At 35 GHz and 20 V, the electronic loss of the structure was about 10 dB. The maximum phase shift of -47 degrees was obtained under these conditions at 25 C.¹

6. Summary

We have fabricated ferroelectric composites of BSTO in bulk ceramics, thick film, and thin film form. The material has exhibited excellent tunable dielectric properties and low electronic loss. It has been demonstrated that these materials are applicable over a wide frequency range (kHz to GHz). The phase shifting microstrip elements of the arrays made from these composites have demonstrated a 360 degrees phase shift with less than 3 dB insertion loss. Coplanar phase shifters have also demonstrated phase shifting abilities that indicate that high frequency operation is also possible. Further work on improvement of the co-planar designs to facilitate low insertion loss at higher frequencies (> 15 GHz) is currently underway.

References

1. R.I. Wolfson, "Phase Shift Device Using Voltage-Controllable Dielectrics", U.S. Patent 5,355,104, Oct. 11, 1994.
2. L.C. Sengupta, S. Stowell, E. Ngo, M.E. O'Day and R. Lancto, "Barium Strontium Titanate and Non-Ferroelectric Oxide Ceramic Composites for Use in Phased Array Antennas," *Integrated Ferroelectrics*, vol 8, pp. 77-88 (1995).
3. E. Ngo, S. Stowell, L.C. Sengupta, M.E. O'Day, and R. Lancto, "Fabrication and Characterization of Barium Strontium Titanate and Non-Ferroelectric Oxide Composites", *Mat. Res. Soc. Symp. Proc.*, vol. 360, pp. 45-50 (1995).
4. L.C. Sengupta, E. Ngo, M.E. O'Day, S. Stowell and R. Lancto, "Fabrication and Characterization of Barium Strontium Titanate and Non-Ferroelectric Oxide Composites for use in Phased Array Antennas," *IEEE Transactions, Proceedings of the International Symposium on the Applications of Ferroelectrics (ISAF)*, (1994) in press.
5. S. Stowell, L.C. Sengupta, E. Ngo, M.E. O'Day and R. Lancto, "Investigation of Aging Effects from Applied Continuous and Pulsed High Voltage Profiles in Ceramic Phase Shifter Materials," *IEEE Transactions, Proceedings of ISAF*, (1994) in press.
6. L.C. Sengupta, S. Stowell, E. Ngo, and M.E. Molongoski, "Thick Film Fabrication Of Ferroelectric Phase Shifter Materials" *J. of Integrated Ferroelectrics*, (1995) in press.
7. M.E. O'Day, L.C. Sengupta, E. Ngo, S. Stowell and R. Lancto, "Processing and Characterization of Functionally Gradient Ceramic Materials," *Proceedings of SPIE Conference, The Int. Soc. for Optical Eng.*, vol 2189, p. 388-399 (1994).
8. C.J. Brennan, *Integrated Ferroelectrics*, vol. 2, pp.73-82 (1992).
9. S. Sengupta, L.C. Sengupta, S. Stowell, D.P. Vijay, and S.B. Desu, "Thin Films Of Novel Ferroelectric Composites" *J. of Integrated Ferroelectrics*, (1995) in press.
10. S. Sengupta, L.C. Sengupta, S. Stowell, D.P. Vijay, and S.B. Desu, "Electrical Characteristics of Barium Strontium Titanate - Oxide Composites" *Mat. Res. Soc. Symp. Proc.*, vol. 360, pp. 413-417 (1995).
11. R.W. Babbitt, T. E. Kosciwa, and W.E. Drach, "Planar Microwave Electro-optic Phase Shifters," *Microwave Jour.*, vol. 35, 63 (1992).

12. W. Drach, T. Koscica, R. Babbitt, L. Sengupta, E. Ngo, S. Stowell, and R. Lancto, " Measurement of Ferroelectrics at Microwave Frequencies and Their Performance in Microwave Devices", *IEEE Transactions* , Proceeding of ISAF (1994) in press.
13. W. Wilber, W. Drach, T. Koscica, L. Sengupta and S. Sengupta, "Fabrication and Performance of Coplanar Ferroelectric Phase Shifter" *J. of Integrated Ferroelectrics* , (1995) in press.

List of Contributors

Tien Pham, Brian Sadler and Manfai Fong
US Army Research Laboratory
Adelphi, MD 20793-1197

Donald Messer
The American University
Washington, DC 20016-8101

Donald M. Maxwell
U. S. Army Medical Research Institute of Chemical Defense
Aberdeen Proving Ground, NM 21010-5425

Kenneth D. Lanclos
Medical College of Georgia
Augusta, GA 30912-2100

Hendrik P. Benschop
TNO Prins Maurits Laboratory
Rijswijk, The Netherlands

V. M. Gold, G. C. Vradis, and J. C. Pearton
US Army ARDEC
AMSTA-AR-AEE-W, Bldg. 3022
Energetics and Warheads Division
Picatinny Arsenal, NJ 07806-5000

L.C. Sengupta, E. Ngo, S. Sengupta and S. Stowell
U. S. Army Research Laboratory
Materials Directorate
APG, MD 21005-5069

Laszlo J. Keeskes
U. S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

Ian W. Hall
University of Delaware
Newark, DE 19716-1501

J. F. Harvey
U. S. Army Research Office
Research Triangle Park, NC 27709-2211

M. B. Steer, J. W. Mink, H.-S. Hwang, and T. W. Nuteson
North Carolina State University
Raleigh, NC 27695-7911

A. C. Paoella
U. S. Army Communications-Electronics Command
Ft. Monmouth, NJ 07703

Michael E. Crenshaw and Charles M. Bowden
Weapons Sciences Directorate, Missile Research, Development, and Engineering Center
U. S. Army Missile Command, Redstone Arsenal, AL 35898 5248

Paul H. Shen, Jagadeesh Pamulapati, Michael Wraback, Weimin Zhou, Monica Taysing-Lara, and Mitra Dutta
U. S. Army Research Laboratory, Physical Science Directorate, AMSRL-PS-PB
Fort Monmouth, NJ 07703-5601

D. A. Reago, Jr. and W. C. Gercken
U. S. Army CECOM Night Vision and Electronic Sensors Directorate
Ft. Belvoir, VA 22060

W. Sturek and S. Ray
U. S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

S. Aliabadi, C. Waters and T. Tezduyar
Aerospace Engineering and Mechanics
Army High Performance Computing Research Center
University of Minnesota
Minneapolis, MN 55415

P. Schihl, W. Bryzik, E. Schwarz, and E. Danielson
U. S. Army Tank-Automotive RDE Center (TARDEC)
Warren, MI 48397-5000

R. C. Sausa and W.R. Anderson
U. S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

G. Singh
University of MD Eastern Shore
Princess Anne, MD 21853-1299

G. W. Lemire
Joint Contact Point Directorate
Dugway Proving Ground, UT 84022-5000

Robert G. Ulrich
U. S. Army Medical Research Institute of Infectious Diseases
Fort Detrick, Frederick, MD 21702-5011

Peter A. Emanuel and James J. Valdes
Research and Technology Directorate, U. S. Army Edgewood Research, Development and
Engineering Center, Aberdeen Proving Grounds, MD 21010

Mohyee E. Eldefrawi
Department of Pharmacology and Experimental Therapeutics, School of Medicine,
University of Maryland, 655 West Baltimore Street, Baltimore, MD 21201

K. K. Choi and W. H. Chang
U. S. Army Research Laboratory, PSD
Fort Monmouth, NJ 07724-5601

C. J. Chen, and D. C. Tsui
Department of Electrical Engineering, Princeton University
Princeton, NJ 08540

LTC Jeff Rabin
U. S. Army Aeromedical Research Laboratory
Fort Rucker, AL 36362-0577

C. A. Biltoft
West Desert Test Center
U. S. Army Dugway Proving Ground
Dugway, UT 84022-5000

E. S. Oran, J. P. Boris, and C. A. Lind
Naval Research Laboratory
Washington, DC 20375

W. J. Mitchell
U. S. EPA Office of Research and Development
Research Triangle Park, NC 27711

Betsy M. Rice, Shannila V. Pai and Cary F. Chabalowski
U. S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

Susan Althoff Gorton, James F. Meyers, and John D. Berry
JRPO-AFDD ATCOM, National Aeronautics, Space Administration
Langley Research Center , Hampton, VA 23681

Herbert A. Leupold
U. S. Army Research Laboratory
Fort Monmouth, NJ 07703-5601

Ming C. Lin
U. S. Army Research Office
Mathematical and Computer Sciences Division
4300 S. Miami Blvd., P.O. Box 12211
Research Triangle Park, NC 27709-2211